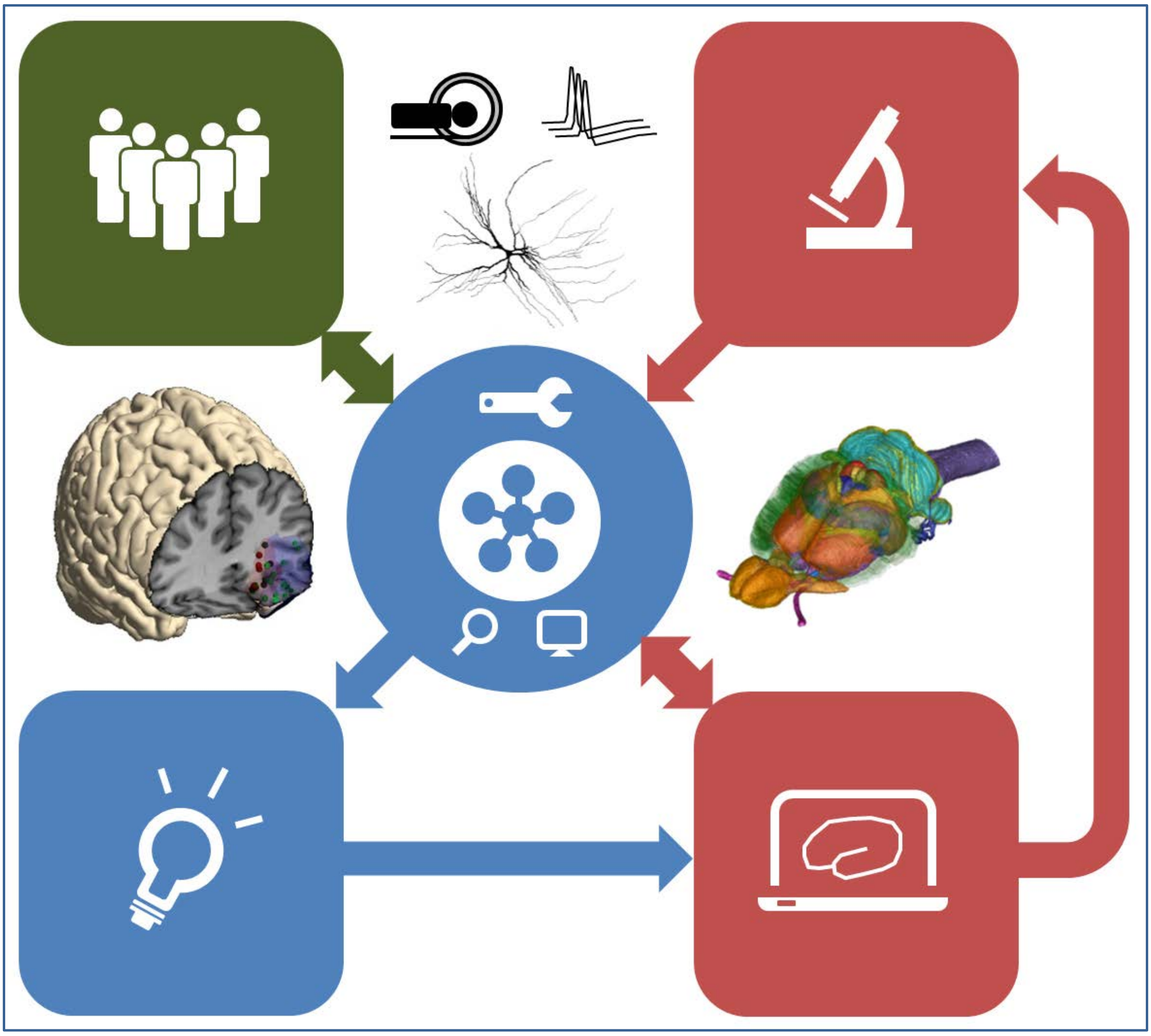

The objective of SP5 - the Neuroinformatics Platform - is to make tools, datasets, and models (blue) discoverable and accessible for researchers in HBP (red) and the wider community (green). SP5 receives data from the experimental neuroscience SPs, organises data in brain atlases, and provides data systems for storage (in collaboration with SP7), retrieval and analysis of research data. SP5 delivers output to modelling and simulation, in turn providing feedback to the experimentalists (red).

Project Number:	720270	Project Title:	Human Brain Project SGA1

Document Title:	SP5 Neuroinformatics Platform - Results for SGA1 Period 1
Document Filename:	D5.8.2 (D31.3 D18 - SGA1 M12) ACCEPTED 180709.docx
Deliverable Number:	SGA1 D5.8.2
Deliverable Type:	Report
Work Package(s):	WP5.1 - WP5.8
Dissemination Level:	PU (= public)
Planned Delivery Date:	SGA1 M12 / 31 March 2017
Actual Delivery Date:	Re-submission to EC November 2017; ACCEPTED 09 Jul 2018

Authors:	SP5: Anastasia AILAMAKI, EPFL (P1); Jan BJAALIE, UIO (P81); Andrew DAVISON. CNRS (P10); Timo DICKSCHEID, JUELICH (P20); Simon EICKHOFF, JUELICH (P20); Pascal FUA, EPFL (P1); Sten GRILLNER, KI (P37); Sonja GRÜN, JUELICH (P20); Michael DENKER, JUELICH P20); Fred HAMPRECHT, UHEI (P47); Anna KRESHUK, UHEI (P47); Pedro LARRAÑAGA, UPM (P68); Trygve LEERGAARD, UIO (P81); Daniel MALLMANN, JUELICH (P20); Jean- François MANGIN, CEA (P11); Jeff MULLER, EPFL (P1); Martin TELEFONT, EPFL (P1); Paul TIESINGA, SKU (P51)
Compiling Editors:	Martha Elisabeth BRIGG, UIO (P81); Sofia ANDERHOLM STRAND, UIO (P81) Roman VOLCHENKOV, UIO (P81); Jan BJAALIE, UIO (P81)
Contributors:	
SciTechCoord Review:	EPFL (P1): Jeff MULLER, Martin TELEFONT, Marie-Elisabeth COLIN UHEI (P47): Martina SCHMALHOLZ, Sabine SCHNEIDER
Editorial Review:	EPFL (P1): Guy WILLIS, Colin McKINNON, Martin O'NEILL

Abstract:	The SP5 work plan was revised up to January 2017. Two of the Work Packages (WPs) in the revised work plan are new: those for the data curation service (WP5.1) and neural activity (WP5.7). The remaining four WPs, covering rodent and human atlasing (WP5.2 and WP5.3), tools developments (WP5.4), and advanced analysis (WP5.6), have undergone various degrees of change in the revised work plan. The most mature parts of SP5 at M12 are the workflows for anchoring of research data to reference atlases and for curation of data in the context of building multi-level atlases of the brain. Together with the general curation services now being established, the new action on neural activity data, and the continuing development of advanced analytics workflows, SP5 will be prepared to receive and organise a broad range of research data during SGA1.
Keywords:	

Table of Contents

List of Figures

List of Tables

1. SP Leader's Overview

1.1 Key Personnel

Subproject Leader: Jan BJAALIE (UIO)

Subproject Deputy Leader: Sten GRILLNER (KI)

Subproject Manager: Sofia ANDERHOLM STRAND (UIO), Roman VOLCHENKOV (UIO)

1.2 Progress update

The work plan for SP5 was extensively revised during the Data Planning Implementation Team (DPIT) process, starting in July 2016 and continuing until the acceptance of the D5.6.1 Resubmission at the end of January 2017.

Two of the Work Packages (WPs) in the revised work plan are new: those for the data curation service (WP5.1) and neural activity (WP5.7). Activities in these WPs have primarily been to recruit personnel, develop detailed plans, initiate work in some tasks, and interact and collaborate with other WPs and SPs in preparation for activities in the remainder of the phase.

The Neuroinformatics Platform (NIP) and infrastructure operations (WP5.5), coordination, and management (WP5.8) are modified in the revised work plan. Several tasks under these WPs are changed considerably, and therefore have just recently started. Also, a new crop of personnel is being or has been recruited, since the research group responsible for major parts of these WPs in the previous SGA1 work plan is no longer participating in the HBP.

The remaining four WPs, covering rodent and human atlasing (WP5.2 and WP5.3), tools developments (WP5.4), and advanced analysis (WP5.6), have had various degrees of changes in the revised work plan. Some Tasks in these WPs (with few changes in the revised work plan) are ongoing and have delivered good progress, whereas others are at an early stage.

Taken together, SP5 is now well prepared for the remainder of SGA1, with an updated work plan and personnel recruited. Collaborations, interactions, and co-developments are ongoing. Routines are in place for regular monitoring of risks. Key performance indicators (KPIs) for all aspects of SP5 have been developed. Overall, SP5 expects to deliver according to the revised work plan at the end of SGA1.

Since M12, more progress has been made on the development and deployment of infrastructure and services, as outlined below and in the Semester report for M13-M18.

HBP data storage was established at CSCS during September, with containers for each task in the HBP work plan. Data flow has been tested with initial storage of data from SP1. Before the CSCS storage became available, SP5 has primarily stored data at JUELICH and UIO using their respective data systems, allowing access to data analysis and data integration in the reference atlases. Progress has been made on developing the KnowledgeGraph (KG), a hybrid semantic-object-graph database system with object schema validation, representational state transfer (REST) application programming interfaces (APIs) and flexible indexing strategies (see 1.2 and 1.3). The achievements in the reporting period represent key steps towards replacing the existing UIO and JUELICH systems with the new HBP service with storage on High-Performance Analytics and Computing (HPAC) servers and search and access through the KG.

The data curation workflow has been established, with three tiers of curation. Basic/general curation (tier 1), intended to cover all data to be ingested, has started with data from SP1, and will continue with data from SP2 and SP3. Curation of location in the brain (tier 2), already well developed, has been further refined in preparation for larger flow of data. Curation of neural activity data (renamed from tier 2 to tier 3 curation) has started with data ingested from SP1. Several ontologies are being aligned to the real-world situation through a direct link between the curation teams and the ontology engineering team, which was ramped up at the end of the reporting period. Selected test datasets from external data repositories, including the Allen Institute mouse resources, have been curated and integrated in the reference atlases.

Development of the infrastructure for feature extraction from images was concentrated on designing and implementing the necessary changes to the ilastik architecture in order to allow for the integration with the HBP tools. The workflows from image series to extracted features, using the HBP reference atlases and tools for registration to atlas, as well as the machine learning capabilities provided by ilastik, have been established and tested. The key remaining aspect for SGA1 is to connect the workflow to the HBP storage, image service, and KG.

Development on the Elephant electrophysiological analysis toolkit has included the addition of several new analytical tools. With the synchronous pattern detection and evaluation (SPADE) method, a major analysis enabled by high-performance computing (HPC) for the detection and evaluation of spatio-temporal patterns in the spike dynamics has been integrated. Rostami *et al*. (2017) is a validation study for Elephant reimplementation of the Unitary Events analysis to detect transient synchronisation of individual neurons. In addition, further work was put into the optimisation and parallelisation of Elephant's analysis of sequences of synchronous events (ASSET) method for the detection of complex sequences in activity data, and the optimised version method is expected to be deployed in Elephant by M24.

1.3 Deviations

The SP5 work plan was last revised in January 2017. Adjustments or corrections have been introduced in a few components as explained in this report. In addition, major delays have occurred in WP5.1 and WP5.5 due to a reorganisation at EPFL that took longer than expected. These are outlined in this section and in the Deviations and corrective actions section of the HBP Semester report for M13-M18.

1.3.1 *Delayed start of basic curation services (WP5.1):*

Reorganisation at EPFL resulted in delayed recruitment of personnel and therefore slower progress than expected under WP5.1 in the reporting period. As a corrective action, resources were transferred to other partners and basic data curation was initiated in a collaborative effort led by UIO. The basic curation team is now working intensively to compensate for the delay.

1.3.2 *KnowledgeGraph not yet operational (WP5.5):*

The recruitment delay has also affected WP5.5. The delay has been mitigated internally at EPFL by merging the efforts of the Blue Brain Project and the Human Brain Project for the development of the KG. Work is now progressing rapidly. The outstanding issues have been cleared and the milestone report (MS5.5.3) has been prepared and is under validation.

Additionally, the KG delay has impacted MS5.4.2 and and MS5.5.2, but with current KG capabilities, the only remaining step is to define appropriate spatial metadata schemas. This work is under way at the time of writing and initial workflow integration is expected to be in use by the end of M21.

The delayed KG is also partly mitigated through the introduction of search functionality in the Data Workbench (the online tool used to harvest metadata, developed under WP5.4), and through the use of the internal systems for organising and managing data at JUELICH and UIO.

1.4 Impact of work done to date

SP5 has made significant progress in establishing workflows for anchoring of research data to reference atlases and for curation of data in the context of building multi-level atlases of the brain. Together with the general curation services now being established, the new action on neural activity data, and the continuing development of advanced analytics workflows, SP5 will be prepared to receive and organise a broad range of research data during the present phase of HBP. With the Collaboratory and the KG hybrid semantic-object-graph database system, WP5.5 is the backbone of the Platform services developed not only in SP5, but throughout the HBP. Enhancements to the Collaboratory improve the usability of this tool for both scientific work and dissemination of HBP Platform advances. The HBP Collaboratory is already having an impact as a key infrastructure component in massively open online courses (MOOCs). The KG is the other key planned outcome for WP5.5 and is a crucial component for findable, accessible, interoperable, and reusable (FAIR) neuroscience data, inside and outside the project. Data accessed through KG queries will be available for viewing and analyses with a suite of shared HBP tools and workflows.

1.5 Priorities for the remainder of the phase

The main priorities for the remainder phase will be to release major parts of the SP5 tools and services, as planned in the DPIT process, through the NIP. See Figure 1. Further details follow.

1. Deployment of DataWorkbench in support of the curation.
2. Tier 1 curation of data submitted to WP5.1 from SPs 1–3.
3. Tier 2 curation of data submitted to WP5.2 and WP5.3.
4. Fully operative tier 3 curation under WP5.7.
5. Evaluate data from external repositories concerning suitability for integration, including validation of accuracy of the registration to reference atlases.
6. Update key ontologies, based on experiences from the curation process.
7. Transfer metadata harvested by curation teams to the Nexus system when this becomes available.
8. Deliver a functional KG metadata data service, prepared also for testing with real spatial metadata, making data curated by HBP findable, interoperable and reusable.
9. Release interactive atlas viewer.
10. Support integration of the KG service into atlasing viewers and initial workflows. This requires initial versions of search applications, curation tools, service client APIs (Python), and integration of performant spatial indexing into the HBP KG Search API.
11. Integrate at least one workflow with image service.
12. Ensure, in collaboration with SP7, that storage integration of the Collaboratory and NIP services supports productive efficient workflows in SP6, SP8, SP9, and SP10.
13. Improve Collaboratory access control list (ACL) models to support more fine-grained separation of write and administration privileges and migration of additional components to a mix of commercial and federated engine for information exchange (FENIX) service providers.
14. Improve Jupyter notebook integration with the planned FENIX services to ensure productive data access for large data files from Jupyter notebooks and other analysis tools.
15. Further strengthen the capacity for SP5 technical and infrastructure coordination, in a continuing close interaction with the HBP Technical Coordinator.

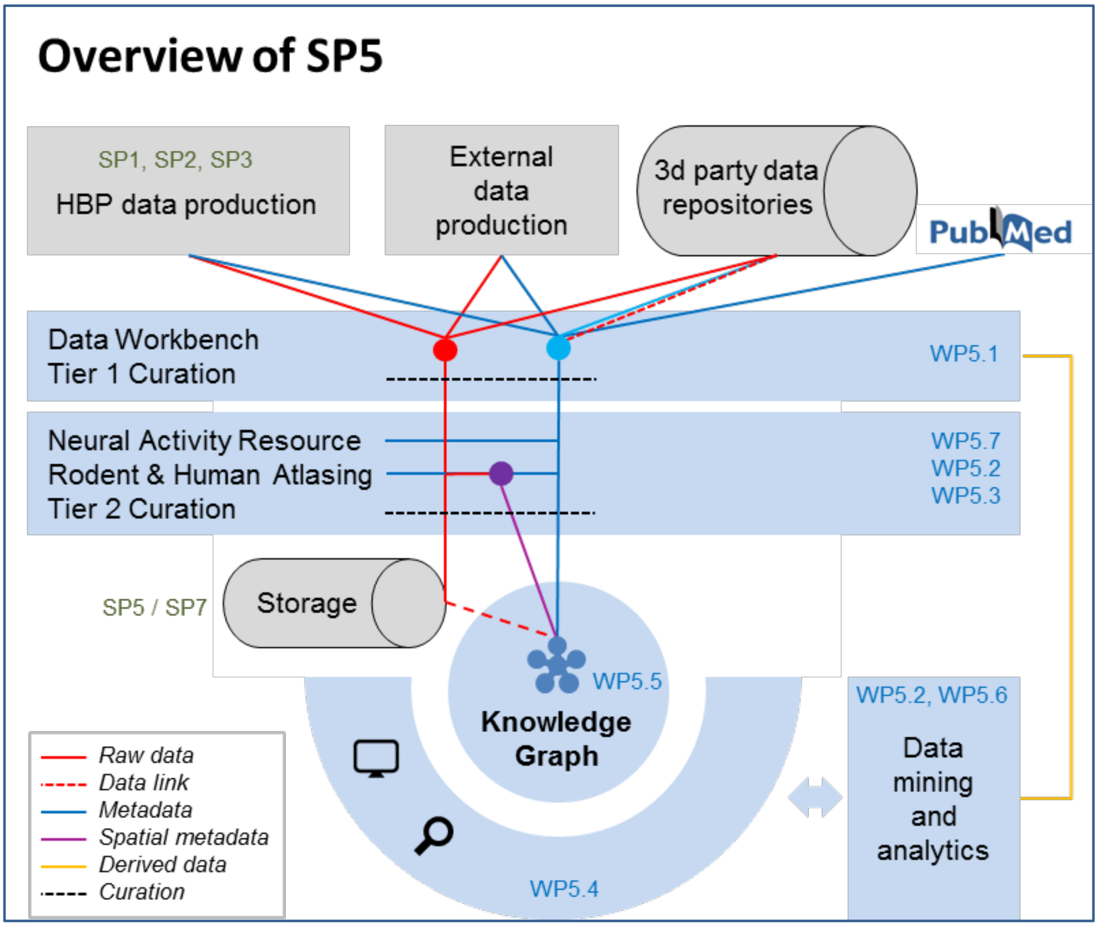

Figure 1: Schematic overview of SP5 activities and data flow.

Data from data-producing SPs in HBP, external research groups, or public third party repositories, are ingested through the HBP Data Workbench before passing through general (Tier 1) and specialised (Tier 2) curation steps. The data are spatially integrated in reference atlases, and metadata are made available for query, mining, and analyses through the HBP Neuroinformatics Platform.

2. WP5.1 Data Curation Support Lab

See Figure 2.

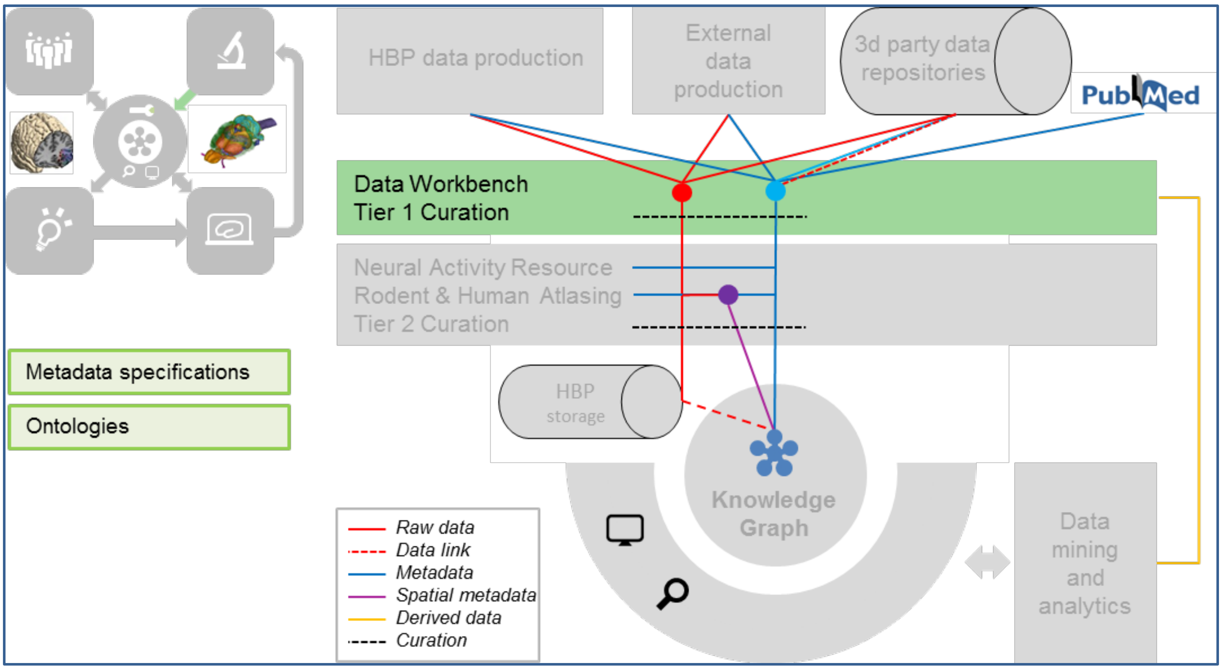

Figure 2: Schematic overview of contributions from WP5.1, highlighted in green.

2.1 Key Personnel

WP Leader: Martin TELEFONT (EPFL)

Task Leaders: Martin TELEFONT (EPFL)

Daniel MALLMANN (JUELICH)

2.2 Work Package Leader's Overview

Hiring in T5.1.1 and T5.1.2 is ongoing and progressing along expected lines. Operational handover of curation work has taken place from the group working on this during the Ramp-Up Phase (RUP). Upgrades of minimal metadata and domain metadata specifications, and ontology development mechanisms is progressing at a good pace. Recently, preparatory work for project scale curation work advanced, and the first results of a first curation push will be noticeable before the M18 report. Timelines for the curation of Tier-1 curation of RUP, and already produced data or models, appears to be holding.

Since this WP is newly established through the revised work plan for SGA2, efforts have so far primarily been at the level of planning and recruiting personnel. It is too early to report on possible deviations from the work plan.

See section 1.3 concerning Deviation and corrective actions related to the late startup of this WP. Following an Amendment to the workplan, the responsibility for this WP is now divided among four partners. The basic curation effort (tier 1 curation, T5.1.1) is currently led by UIO.

2.3 Priorities for the remainder of the phase

In the remainder of the phase, we will organise the work in iterative waves of curation and subsequent tool improvements of interface and APIs (with T5.4.1).

2.4 Milestones

See Table 1.

Table 1: Milestones for WP5.1 — Data Curation Support Lab

MS No.	Milestone Name	Lead Partner	Task(s) involved	Month Due	Month of Completion	Comments
MS5.1.1	List of KPIs for internal and external user adoption of Products/services offered by WP5.1	1 (EPFL)	T5.1.1	M10	M10	Revised and finalised by SP leader. Approved by task/WP leaders. Minor adjustments of KPIs for internal SP use will be considered based on accumulated experience during the project period. A pre-verification version of this Milestone was delivered to the European Commission together with the D5.6.1 Resubmission in January 2017. The full list of KPIs is available for the HBP SGA1 consortium through the Milestone Report Repository and can also be found in Appendix 1. In this report, we include the M18 report on KPIs, with comments on adjustments of targets for SGA1

2.5 T5.1.1 - Data and Parameter Workbench Support and Curation Team

GENERAL NOTE CONCERNING PRODUCTS AND COMPONENTS: Please note that in this report, the term 'Component' is used, rather than the term 'Product' which is the term used in the D5.6.1 Resubmission.

2.5.1 *Key Personnel*

Task Leader: Martin TELEFONT (EPFL)

2.5.2 *SGA1 DoA Goals*

- Ensure all RUP/SGA1 data are discoverable in services provided by T5.4.1.
- Assist HBP researchers in adopting tools developed in T5.4.1.
- Assist HBP researchers in adopting enrichment workflows developed in T5.4.1.
- Organise, verify and improve metadata enrichments provided by SPs 1–10 and external researchers, to improve discoverability of data and parameters.

2.5.3 *Component Progress*

NOTE: Component 1-1 from the D5.6.1 Resubmission has been divided into two components, a and b, to distinguish data and models from RUP and SGA1.

2.5.3.1 Component 1-1a Meta-Data used to enrich RUP data and models

Description of Component: RUP Data and Models will be enriched so as to make them discoverable via the NIP.

Progress on Component: Tier 1 curation has started with data from SP1.

2.5.3.2 Component 1- 1b Meta-Data used to enrich SGA1 data and models

Description of Component: SGA1 Data and Models will be enriched so as to make them discoverable via the NIP.

Progress on Component: Tier 1 curation has started with data from SP1.

2.5.3.3 Component 2-2 Living report on adoption of T5.1.1 tools by HBP and external Researchers

Description of Component: This report will provide an overview of how the outreach and training Component of this Task fared.

Progress on Component: A reporting system monitoring the ongoing curation process has been implemented. A snapshot showing the functionalities is shown in Figure 3.

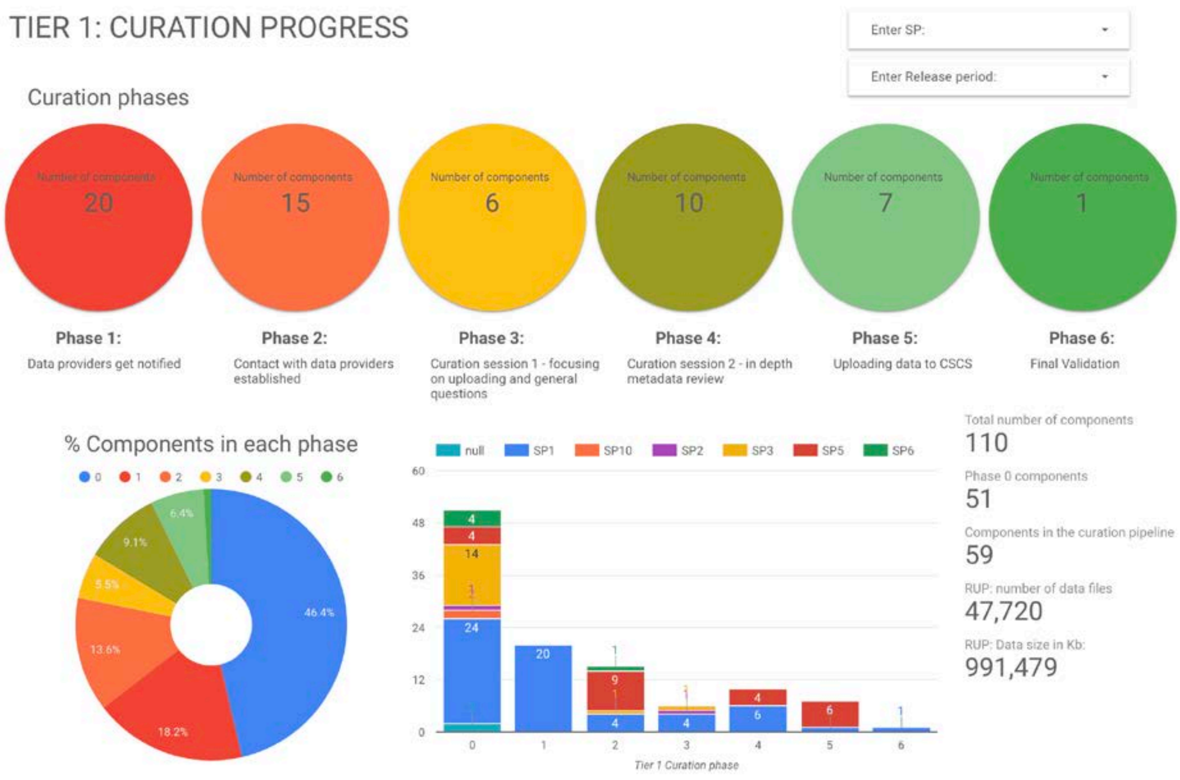

Figure 3: Monitoring of the tier 1 curation process. Each component in the workplan is followed through the different steps, from notification of curation startup to final validation of metadata.

2.6 T5.1.2 - Ontologies for the integration of Models in the KnowledgeGraph

2.6.1 *Key Personnel*

Task Leader: Martin TELEFONT (EPFL)

2.6.2 *SGA1 DoA Goals*

Knowledge engineering of the ontologies required to support the data-driven model building.

2.6.3 *Component Progress*

2.6.3.1 Component 3-1 Identification of HBP Users' Use Cases

Description of Component: In order to build scientifically relevant ontologies, we will need to interview HBP users in order to understand which Use Cases our ontologies need to support. We will document these Use Cases, as well as the resulting ontology architecture.

Progress on Component: Interaction with SP6 users has taken place and needs were taken into account for the first wave of curation push preparatory work.

2.6.3.2 Component 4-2 Ontology engineering

Description of Component: ontology for data/model discoverability annotation.

Progress on Component: Background work was done by the WP Leader (see comments in section 2.2). Team working on ontology engineering now active.

2.7 T5.1.3 - Large-Scale Federated Data Accessibility Services

2.7.1 *Key Personnel*

Task Leader: Daniel MALLMANN (JUELICH)

2.7.2 *SGA1 DoA Goals*

Leveraging HBP storage services in SP7 and the Collaboratory.

2.7.3 *Component Progress*

2.7.3.1 Component 6-1 Support for data upload and download

Description of Component: Supporting users to upload data to a repository in the federated data infrastructure or to download it from such a repository.

Co-Design Project (CDP) to which Component contributes: This component has so far contributed to CDP3, Products 3 and 4 (CDP3-P3 and CDP3-P4).

Progress on Component: Users that want to upload data to a repository need to have an HBP HPAC Platform (bp0x) account. For the download of data from such a repository, a REST API is implemented that allows for publishing data on a uniform resource locator (URL). There are Use Cases that need to restrict download access to users of the Collaboratory (or a user subgroup, e.g. SP5 users) and discussions on how to support this are going on in SP7. Workflows for data uploading to JUELICH's General Parallel File System (GPFS) system have been successfully tested, and since September, dedicated Swift Object Storage for HBP data is available at CSCS. The storage can be easily grown in incremental steps according to changing storage needs. The flow of data from data providers through the curation teams to the HBP storage is now operational, with usability improvements to be introduced over the coming months.

2.7.3.2 Component 7-2 Support for data transfers

Description of Component: Supporting users to transfer data between an archive and an active repository in the federated data infrastructure.

Progress on Component: No support requests for data transfer between an archive and an active repository so far.

2.7.3.3 Component 8-3 Support for data searches

Description of Component: Supporting users to search for datasets stored in the federated data infrastructure from the Collaboratory.

CDP to which Component contributes: The component has so far contributed to CDP3-P3 and CDP3-P4.

Progress on Component: We had no support requests for data searches so far, but have focused on the data upload and download as the initial service.

3. WP5.2 Multi-level Atlas of the Rodent Brain

See Figure 4.

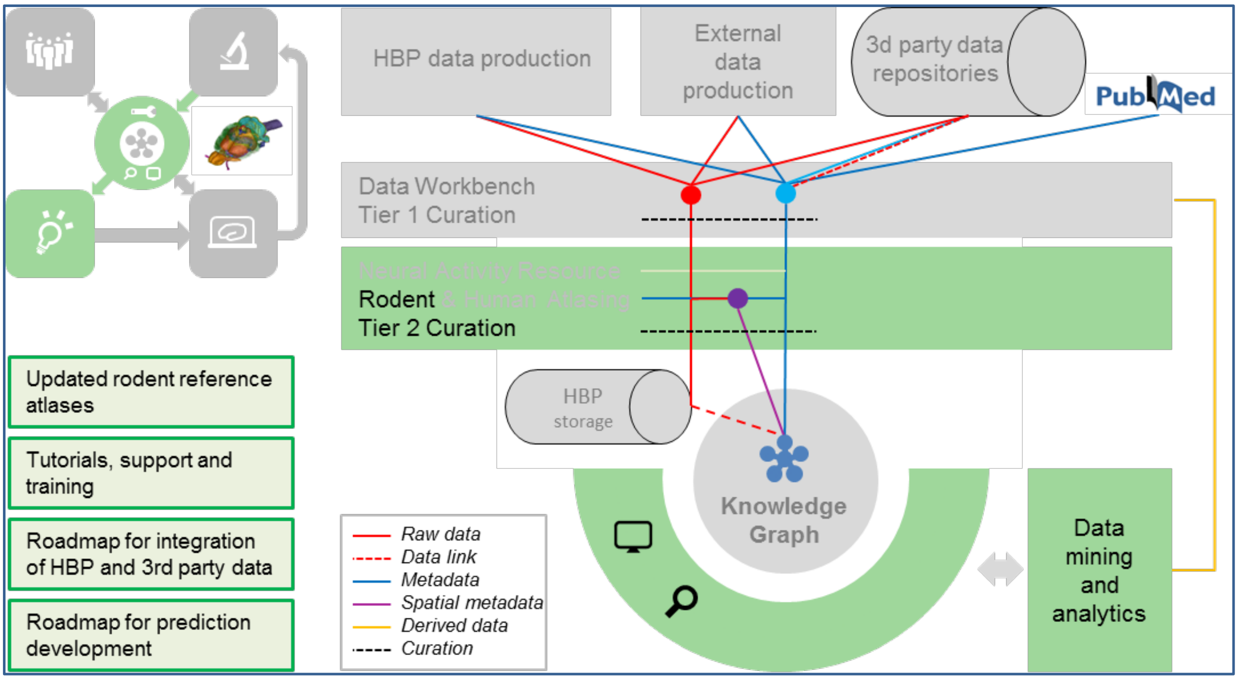

Figure 4: Schematic overview of contributions from WP5.2, highlighted in green.

3.1 Key Personnel

WP Leader: Jan BJAALIE (UIO)

Task Leaders: Trygve B. LEERGAARD (UIO)

Jan BJAALIE (UIO)

Paul TIESINGA (SKU, JUELICH)

3.2 Work Package Leader's Overview

Curation team and workflow for anchoring of images to atlas have been established: 1) training of personnel has been completed; 2) tools have been extensively tested and used; and 3) regular contact with data-producing researchers in SP1 and SP3 (partly in context of CDP1) has been established.

Atlas maintenance, including additions of structures to reference atlases, is ahead of schedule.

Anchoring of HBP data to reference atlases is proceeding as planned.

Interactions with the WP5.1 Tier 1 curation team at EPFL started in M12 and has continued with the new team (see report under WP5.1).

Work in WP5.2 during the first 12 months was adjusted to the fact that the KG and FENIX storage infrastructure were not yet operational. Temporary solutions were used for storing data and metadata in a fashion allowing easy migration to the final systems. HBP storage is now in place and workflows for data uploading have been established (see also WP5.1 and WP5.5). An update on the KG development is provided under WP5.5.

Several use-cases with anchoring and curation of SP1 and SP3 data confirm that the teams, tools, and procedures are functioning as planned.

3.3 Priorities for the remainder of the phase

The priority for the remainder of SGA1 will be to anchor data to reference atlases at the performance level indicated by the relevant KPIs. Anchoring of data will be curated by the WP5.2 team, following either anchoring by the researchers producing the data or anchoring performed by the WP5.2 team. Data are currently transferred to the new HBP storage and metadata will be submitted to the KG when services become available. Complete workflows from submission of image data to storage and retrieval will be established and tested. In the period up to the HBP Summit (October 2017), collaborative efforts focused on interaction with data producers (SP1/SP3), tier 1 curation teams (WP5.1), tools and system developers (WP5.4, WP5.5), data analysis specialists (WP5.6), and modellers (SP4, SP6). In the last part of SGA1, the focus will shift towards increasing the amount of interaction with the broader scientific community at the level of providing training in the use of tools and access for selected users to workflows and services.

3.4 Milestones

See Table 2.

Table 2: Milestones for WP5.2 — Multi-level Atlas of the Rodent Brain

MS No.	Milestone Name	Lead Partner	Task(s) involved	Month Due	Month of Completion	Comments
MS5.2.1	KPIs for rodent atlas user adoption	81 (UIO)	T5.2.1-5	M10	M10	Revised and finalised by SP leader. Approved by task/WP leaders. Minor adjustments of KPIs for internal SP use will be considered based on accumulated experience during the project period. A pre-verification version of this Milestone was delivered to the European Commission together with the D5.6.1 Resubmission in January 2017. The full list of KPIs is available for the HBP SGA1 consortium through the Milestone Report Repository and can also be found in Appendix 1. In this report, we include the M18 report on KPIs, with comments on adjustments of targets for SGA1
MS5.2.2	Two-dimensional (2D) and three-dimensional (3D) image data from HBP Partners spatially anchored to atlas with metadata in KnowledgeGraph	81 (UIO)	T5.2.2	M12	M12 Partly	Components providing input to Milestone: Data components of SP1 and SP3, CDP1. This Milestone has two parts: anchoring to atlas and metadata in KG. The first part has been achieved and metadata have been prepared. The second part will be achieved when the KG is operative. Forecast achievement date for metadata in KG: 1 October 2017. Forecast verification date: 1 November 2017. Successful anchoring of image data and storage of metadata in KnowledgeGraph will be verified by use of NIP query tools. Verification will be performed by data producers and UIO and EPFL curation teams and approved by the WP leader. All RUP and SGA1 data available for Tier 2 curation have been spatially anchored to atlas and ensuing spatial metadata prepared for storage in KnowledgeGraph. Once the KnowledgeGraph and Tier 1

						curation are operative, the spatial metadata will be merged with other experimental metadata and entered in the KnowledgeGraph. The full M12 report on this Milestone is available for the HBP SGA1 consortium through the Milestone Report Repository. This milestone has now been achieved to the level of having metadata prepared for entry into the KnowledgeGraph. The report is available in EMDESK https://emdesk.humanbrainproject.eu/shared/5a098c6a7a74a-0c808687ca1975cf98a4ceb5005666e8 Examples of data anchored to atlas space can be found in the collab "Data anchored to reference atlases": https://collab.humanbrainproject.eu/#/collab/5366/nav/41799
MS5.2.3	SGA1 and SGA2 Roadmap for integration of HBP and third party rodent data into the rodent atlases. Validated by the CDPs and SP6 (T5.2.2)	81 (UIO)	T5.2.2 T5.2.3 T5.2.4	M12	M12	This roadmap is critical for planning of SGA2 efforts, and input from partners outside SP5 will be required. Several external data repositories have been evaluated, and a limited number of available example data sets downloaded and anchored to reference atlas space [see Progress comments to T5.2.3 (section 3.7.3) and T5.2.4 (section 3.8.3)]. The Roadmap has been verified by the curation team, evaluated and verified by task leader, and approved by WP and Task leaders A full report on this Milestone is available for the HBP SGA1 consortium through the Milestone Report Repository. https://emdesk.humanbrainproject.eu/shared/59a963a46064d-598613c8b793b5f475c97e3a204aa8b3. At the time of submission of D5.8.2, milestones were uploaded in an HBP repository. Unfortunately, reviewers did not have access to this repository due to technical limitations
MS5.2.4	Prioritised roadmap of properties targeted for prediction development. Validated by the CDPs and SP6	51 (SKU)	T5.2.5	M12	M12	The Roadmap proposal has been commented and approved by WP and Task leaders. The full report is available for the HBP SGA1 consortium through the Milestone Report Repository.

						https://emdesk.humanbrainproject.eu/shared/59a963d4c3b4d-85644e0f442872e37ffa2092d4674e98. At the time of submission of D5.8.2, milestones were uploaded in an HBP repository. Unfortunately, reviewers did not have access to this repository due to technical limitations

3.5 T5.2.1 - Maintenance of Rodent Atlases

3.5.1 *Key Personnel*

Task Leader: Trygve LEERGAARD (UIO)

3.5.2 *SGA1 DoA Goals*

Maintain and update the content and functionality of the rodent reference atlases, including

- updating of image templates used for the 3D reference atlas;
- management of image templates and delineation files constituting the reference atlases;
- validating, integrating and sharing new or adjusted delineations;
- compiling downloadable files in different formats.

Efforts to update the atlases include revision and expansion of anatomical delineations by interpretation of image templates enriched with spatially registered microscopic data, and sharing of customised atlas versions with granularity tailored for different analytical purposes.

3.5.3 *Component Progress*

3.5.3.1 Component 9-1 Waxholm Space

Completed in RUP.

3.5.3.2 Component 10-2 Waxholm Space rat brain reference atlas, v1.0

Completed in RUP.

3.5.3.3 Component 11-3 Waxholm Space rat brain reference atlas, v1.0.1

Completed in RUP.

3.5.3.4 Component 12-4 Waxholm Space rat brain reference atlas, v2.0

Completed in RUP.

3.5.3.5 Component 13-5 Enriched Waxholm Space rat brain reference atlas: PLI/M2 data

Completed in RUP.

3.5.3.6 Component 14-6 Atlas of rat brain cyto-and myeloarchitecture

Completed in RUP.

3.5.3.7 Component 15-7 Whole brain connectivity atlas of the rat

Completed in RUP.

3.5.3.8 Component 16-8 Volumetric Allen mouse brain reference atlas

Completed in RUP.

3.5.3.9 Component 17-9 Update of Allen mouse brain reference atlas

Completed in RUP.

3.5.3.10 Component 18-10 Three-plane high resolution architectonic atlas of the rat hippocampal region

Completed in RUP.

3.5.3.11 Component 19-11 Standardised anatomical landmarks for registration of whole brain imaging datasets to mouse brain Waxholm Space

Completed in RUP.

3.5.3.12 Component 20-12 Tg-Arc-Swe atlas

Completed in RUP.

3.5.3.13 Component 21-13 Neurotransporter atlas: GLT1

Completed in RUP.

3.5.3.14 Component 22-14 Neurotransporter atlas: VGLUT3

Completed in RUP.

3.5.3.15 Component 23-15 tTA mouse brain atlas: Nop

Completed in RUP.

3.5.3.16 Component 24-16 tTA mouse brain atlas: PrP

Completed in RUP.

3.5.3.17 Component 25-17 tTA mouse brain atlas: PrP-CamKII

Completed in RUP.

3.5.3.18 Component 26-18 Database on rat cerebro-cerebellar connectivity

Completed in RUP.

3.5.3.19 Component 27-19 Customised versions of Allen mouse brain atlas tailored for different analyses

Description of Component: The full version of the Allen mouse brain reference atlas contains a very large number of small structures. For many types of analyses, this level of granularity is inappropriate. To facilitate semi-quantitative analysis of spatial distributions of labelled markers in images that have been anchored to the Allen mouse brain reference atlas, we will create several custom versions of the atlas with different granularity (number of structures). These custom versions will be bundled in the QuickNII tool.

Progress on Component: Ongoing. The Allen mouse brain reference atlas (v2) contains 738 delineations, many of which are very small. Three different versions of the Allen mouse brain atlas delineations have now been prepared and bundled in the QuickNII tool (see section 3.5.3.20). The “Rainbow” version uses randomised colouring in subregions to increase the visibility of small regions (i.e. cortical layers). Two “Reduced” versions have been created in which the number of atlas delineations has been selectively pruned to contain ~200-300 delineations be more suitable for efficient semi-quantitative analyses of labelled objects in regions of interest in a collection of RUP data (RUP products 23-15, 24-16, 25-17). These atlas versions are disseminated together with the QuickNII tool (T5.4.2). Similar versions are in preparation for the most recent Allen mouse brain reference atlas (v3).

3.5.3.20 Component 28-20 Allen mouse brain reference atlas with white matter structures parcellated

Description of Component: In the Allen mouse brain reference atlas, white matter regions are not segmented as individual regions. This is a limitation for several visualisation and analytic purposes. We will therefore create an updated version of the atlas where the white matter is parcellated into separate smaller structures.

Progress on Component: Completed. Segmentations of mouse brain white matter regions are included in v3 of the Allen mouse brain reference atlas, and this version has now been incorporated in the tools used for anchoring to atlas and subsequent visualisation of images in atlas space. Since the new version of the Allen mouse brain atlas is defined in a new population average template, which is spatially incongruent with the earlier version of the atlas, mouse brain data already registered in the atlas template v2 were re-aligned with the new atlas. This change in template prompted the need for a coordinate transformation tool, now proposed as a component for SGA2 (https://project-lifecycle.herokuapp.com/component/2279/).

3.5.3.21 Component 29-21 Waxholm Space rat brain reference atlas enriched with additional receptor data

Description of Component: Enrichment of the Waxholm Space (WHS) rat brain reference atlas with brain-wide data on the spatial distribution of excitatory, inhibitory, and modulatory receptors, expanding on the results from RUP efforts. These data will be used as strategic data for modelling purposes and to drive efforts to further delineate the WHS reference atlas.

Progress on Component: Ongoing. A first set of muscarinergic receptor data were registered in the rat brain reference template (RUP product 13-5). Atlas delineation based on these data has commenced, but due to unforeseen technical challenges, additional receptor data are not ready. The primary motivation for this Component has been to drive the further delineation of the reference atlas with the addition of new data types. Thus, as a partial replacement and additional substrate for continued atlas delineation, a comprehensive collection of coronal, sagittal, and horizontally cut histological sections stained by Timm's sulfide silver method will be anchored to the Waxholm template. Timm's staining reveals distribution of metals in the brain, including zinc-positive boutons, and is highly useful for identification of patterns of lamination in brain tissues, particularly in the hippocampus. The material has now been digitised and collections of microscopic images prepared for spatial registration in the atlas.

3.5.3.22 Component 30-22 Expanded and improved Waxholm Space rat brain reference atlas (with additional and corrected structure delineations) based on registered multimodal data

Description of Component: The Waxholm Space rat brain reference atlas has been delineated in several steps. Enrichment of the reference atlas by spatial registration of different types of experimental data (see also T5.1.2) provides new substrates for further and more refined delineations of additional brain structures and systems. Utilising new data registered in the atlas space, we will update and expand this reference atlas resource. The WHS rat brain atlas is already widely used (>41,000 downloads) and improved delineations will have great value for the community.

Progress on Component: Nearly completed. Efforts with the expansion of the Waxholm Space rat brain atlas (v2.0, RUP product 12-4) has been focused on the auditory system, as a follow-up of the cytoarchitectonic parcellation of the human auditory cortex achieved in the human BigBrain atlas by JUELICH. See Figure 5. Structures in the auditory system (brainstem, midbrain, and cortical regions) have been identified in the high-resolution magnetic resonance imaging/diffusion tensor imaging (MRI/DTI) Waxholm Space rat brain atlas template. Detailed delineations have been made possible by the use of RUP data registered in the atlas template, including multiplane microscopic images showing normal myelo- and cytoarchitecture (RUP products 14-6, 18-10) and 3D polarised imaging data (RUP product 13-5). More than 30 new brain structures have been added, and delineations of several other midbrain and hindbrain structures adjusted or corrected. Work is under way to document the new delineations and prepare the release of the next version.

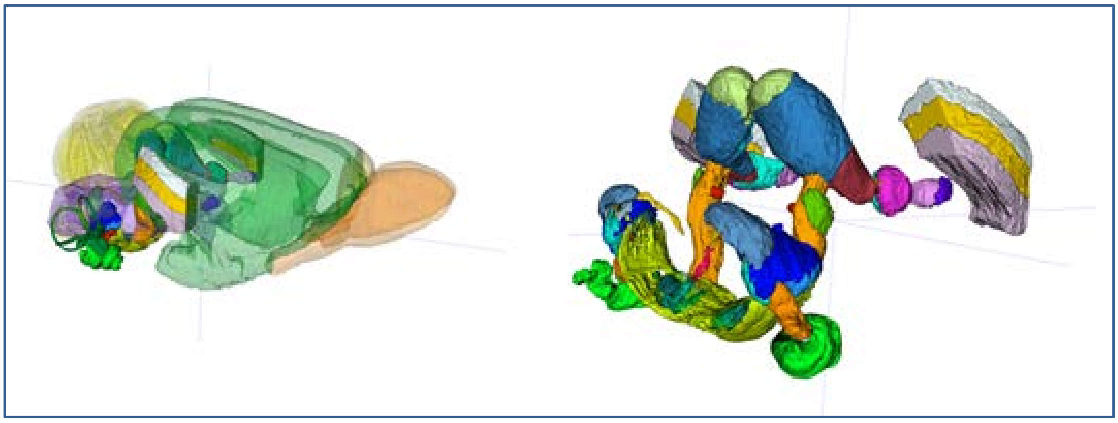

Figure 5: Updated Waxholm Space atlas of the rat brain, with additional delineations of the auditory system.

3.6 T5.2.2 - Curation of Rodent Atlas Data

3.6.1 *Key Personnel*

Task Leader: Trygve LEERGAARD (UIO)

3.6.2 *SGA1 DoA Goals*

Tier 2 curation of anatomical location parameters assigned to experimental data, delivering spatial metadata to the KG.

This Task will:

- provide efficient workflows, tutorials, and training for data-producing researchers (SPs 1-3) performing anchoring of data;
- optimise procedures for assignment of spatial location to the heterogeneous rodent brain image data produced within the HBP (spanning molecular, cellular, and system levels), captured using a wide range of methods;
- define minimum metadata requirements needed to ensure accurate determination of data locations in atlas space;
- validate and approve quality of spatial anchoring prior to submission of spatial metadata to the KG.

3.6.3 *Component Progress*

3.6.3.1 Product 31-1 Tutorials, training and supervision in assignment of spatial metadata

Description of Component: The QuickNII tool (see T.5.4.2) is quite mature and a workflow for spatially defining large numbers of serial image data to reference atlas space has been tested and optimised. We will produce a tutorial and course material for providing practical training in the use of the tool and associated procedure. The UIO team will train users in hands-on courses, as well as remotely by email and teleconference contact. Data produced in the RUP (see T5.2.1) will be used for training purposes. Support will be given to researchers using the tools and procedures at HBP meetings, by phone, email, and teleconference.

Progress on Component: Ongoing. This component is a fully operative training service for HBP and external researchers. Workflow and associated tutorials developed during the RUP were used to train the newly recruited UIO curation team. Accumulated experiences were used to improve the tool, workflow, and training material. By SGA1 M6, several practical workshops had been given, including a workshop at the HBP 2016 Summit in Florence. Software and course materials were also provided to neuroscientists giving a course at the

Howard Florey institute in Australia. So far, 35 users (HBP and external) have been trained by the SP5 team in spatial registration of images to reference atlases using QuickNII. Information about the tool and workflow has been disseminated at HBP and institutional meetings, and further dissemination is planned, including presentations at international meetings and as an original publication documenting the tool and workflow.

3.6.3.2 Component 32-2 Validation and approval of spatial metadata before final entry into KnowledgeGraph

Description of Component: Spatial metadata assigned to experimental data will be reviewed by a team of expert curators who will evaluate spatial accuracy and completeness, and approve data for ingestion in the KG.

Progress on Component: Ongoing. This Component concerns the fully operative Tier 2 curation services. See Figure 6. So far (since the startup of HBP) >16,000 images from >90 rodent brains have been anchored to reference atlas space and validated providing spatial metadata ready for ingestion into the KG. All currently available image data have been anchored, and nearly all of these have been curated. The curation team has established direct contact with several SP1 and SP3 groups (Francesco PAVONE, Javier DEFELIPE, Pedro LARRAÑAGA, Francisco CLASCÁ, Cyriel PENNARTZ, Viktor JIRSA) to anchor and curate their experimental rodent brain data. See Figure 7. The curation team will closely collaborate with the Tier 1 curation team at EPFL to curate all accumulated RUP and SGA1 rodent brain data (WP5.1). Spatial anchoring and curation services have also been provided to external (non-HBP) laboratories, including Edvard I. MOSER (NTNU, Norway), Menno P. WITTER (NTNU, Norway), and Koen A. VERVAEKE (UIO, Norway).

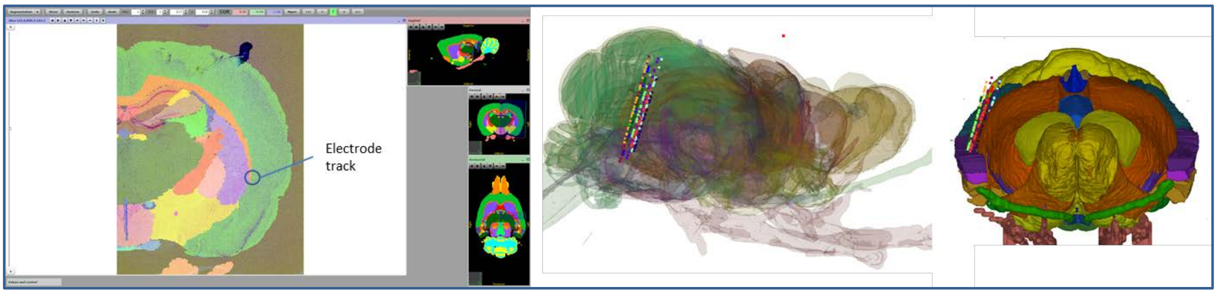

Figure 6: Graphical user interface of the QuickNII tool (left) showing an image of a histological section from the PENNARTZ laboratory, used to identify location of electrode tracks.

The MeshView tool (middle and right) is used to visualise the locations of multiple electrode tracks in the Waxholm Space atlas of the rat brain. The recording sites were identified as located in the perirhinal areas 35 and 36.

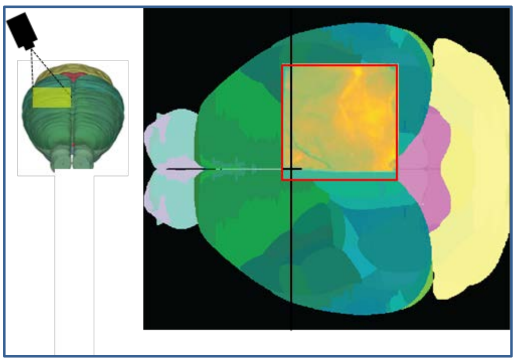

Figure 7: Cortical activity map (calcium imaging data), data from laboratories of PAVONE and JIRSA, mapped on to the Allen mouse brain atlas with use of the QuickNII and MeshView tools.

3.6.3.3 Component 33-3 Optimised procedure for anchoring of 2D image data to reference atlas

Description of Component: As optimised versions of the QuickNII tool for spatial registration are released for HBP and public use, data registration procedures, and associated tutorials will be updated and optimised, also taking into account accumulated user experiences.

Progress on Component: Completed. The pilot procedure established during RUP has been developed and extensively tested. So far (since the startup of HBP) >16,000 images from >90 rodent brains have been anchored to reference atlas space. Following an agile development process, the QuickNII tool has been stepwise improved (T5.3.2). Monitoring of the number of hours used for different workflow steps show that the majority of work is spent on preprocessing steps needed to organise the image material (including correction of serial order and image orientation), while anchoring takes less time. This demonstrates the need to establish best practice recommendations for preparation and documentation of different types of experimental neuroscience image data. Development of such recommendations is ongoing and will be disseminated within the HBP. A more complete overview to cover a broad range of image data types requires further effort and has therefore been proposed as a Component for SGA2. Accumulated experiences have been used to update procedure descriptions and tutorials (see below), and suggestions for further improvements of tool functionality have been communicated to the software developer team (T5.4.2).

3.6.3.4 Component 34- 4 Procedure for anchoring of 3D image data to reference atlas

Description of Component: This procedure will build on the QuickNII tool and procedures for anchoring 2D data to reference atlas space, and be adapted for handling of volumetric image data as series of pre-aligned 2D images.

Progress on Component: Completed. Several 3D image volumes have been successfully anchored to reference atlas space, including: 1) serial two-photon tomography volumes downloaded from the Allen mouse brain connectivity atlas; 2) 3D histology reconstructions showing rat brain connectivity (RUP product 15-7); and 3) volumetric light sheet microscopy data (T1.3.1/CDP1). See Figure 8. Procedure descriptions have been prepared, and suggestions for improved tool functionality have been communicated to the software developer team (T5.4.2).

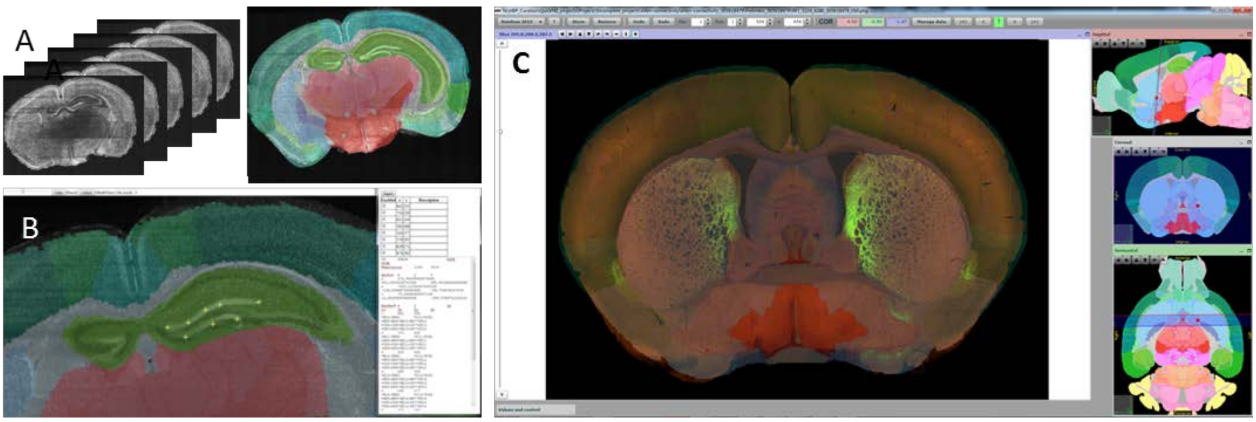

Figure 8: A-B: Light sheet data (A) from the PAVONE laboratory, propidium iodide staining, mapped into the Allen mouse brain atlas using the QuickNII tool.

Further analysis of the images is exemplified in B. The LocaliZoom tool is here used to extract atlas coordinates of identified elements in the experimental material. C: Whole brain connectivity data downloaded from the Allen Institute are shown in atlas space using the QuickNII tool.

3.6.3.5 Component 35-5 Procedure for non-linear warping of 2D image data to reference atlas

Description of Component: This procedure will describe a methodology for non-linear warping of 2D image data to reference atlas space, as a follow-up after anchoring of images using QuickNII.

Progress on Component: Ongoing. A first version of the software tool has been developed and tested (T5.4.2). The tool allows 2D images that have already been anchored linearly to atlas space with the QuickNII tool to be non-linearly deformed and displayed in native atlas space. The next steps are to identify transformation parameters for different types of 2D image data, optimise the procedure, and prepare a user interface for non-experts. Once the procedure is operative and tested, workflow descriptions and tutorials will be prepared.

3.6.3.6 Component 36-6 Procedure for non-linear warping of 3D image data to reference atlas

Description of Component: Methodologies for warping of image volumes to reference atlas space exist, but require specialised tools and expertise. We will first establish procedures for non-linear warping of different data types (T5.1.2), and secondly develop these for use by non-experts.

Progress on Component: Ongoing. The software tool to be used is under development (T5.4.2). A selection of data to be used has been collected: 1) rat brain MRI/DTI volumes acquired with the same parameters as the Waxholm Space template;[1] 2) the Waxholm Space mouse brain template;[2] 3) serial two-photon tomography volumes downloaded from the Allen Mouse brain connectivity atlas; 4) 3D histology reconstructions showing rat brain connectivity (RUP product 15-7); and 5) volumetric light sheet microscopy data (T1.3.1/CDP1). The next steps are to evaluate different parameters for feature-based non-linear registration of these volumes to rat or mouse atlas templates and to assess to what extent a first step affine registration of volumes using the QuickNII tool can improve the registration procedure and result. Once the procedure is operative and tested, workflow descriptions and tutorials will be prepared.

3.6.3.7 Component 37-7 Curation of semantic spatial metadata delivered in T5.4.1

Description of Component: Here a team of experts in assignment of anatomical location will evaluate the granularity, accuracy and completeness of semantic spatial metadata assigned to experimental data delivered in T5.1.1, prior to ingestion in the KG.

Progress on Component: Ongoing. Tools for metadata curation are not yet available, awaiting recruitment of personnel to these tasks (EPFL, Telefont). Meanwhile, a procedure for assessing the completeness and accuracy of spatial metadata has been established. The procedure is used to describe the anatomical location of different types of experimental neuroscience data, and has been validated on a collection of 120 publications sampled from PubMed. This investigation demonstrates that the majority of experimental publications contain highly insufficient descriptions and documentation of the spatial location of presented observations. Based on these results, a list of recommended spatial metadata and a pilot metadatabase has been created in collaboration with the Tier 1 curation team (WP5.1), and a new component defining better practice recommendations for documentation of spatial metadata has been suggested for SGA2.

3.7 T5.2.3 – Harvesting and Curation of Strategic Metadata from Existing Third Party Data Repositories

3.7.1 *Key Personnel*

Task Leader: Jan BJAALIE (UIO)

3.7.2 *SGA1 DoA Goals*

Make a selection of publicly available datasets hosted in external neuroinformatics data repositories discoverable through NIP.

Efforts are focused on identifying and evaluating data collections that have high relevance for predictive analytical and modelling Tasks, as well as establishing procedures for harvesting data from important repositories, and then validate them by harvesting and curating 5-10 complete data sets from external repositories.

3.7.3 *Component Progress*

3.7.3.1 Component 38-1 Metadata and links to Allen Institute mouse brain data repositories

Description of Component: Strategic mouse brain gene-expression and connectivity data shared in Allen Institute repositories will be selected for ingestion into the KG in collaboration with researchers in SP4 and SP6. Required metadata will be entered using tools and forms provided in T5.4.1, and spatial metadata will be added using established tools (T5.4.2) and procedures (T5.2.2), curated, and released to the KG together with links pointing to the native data stored at the Allen Institute (https://www.alleninstitute.org/).

Progress on Component: Ongoing. The Allen Institute provides access to several large repositories containing numerous microscopic images that can be inspected in web-microscopy viewers with several interactive functions or downloaded using APIs. The mouse brain image collections include data from >2,500 tract-tracing experiments, >20,000 *in-situ* hybridisation experiments, electrophysiological recordings and 3D reconstructions from >844 visual cortex cells, 112 functional imaging experiments from the visual cortex, and ~14,000 *in situ* hybridisation experiments showing gene expression in the developing mouse brain. The image data are either serial two-photon tomography data or digitised histological serial sections. All data have been mapped to the Allen Mouse brain atlas template using automated methods.

So far, six different complete datasets have been downloaded using automatic APIs and successfully anchored using QuickNII. The next step is to evaluate the accuracy of spatial registration in the atlas space, and to determine whether the available STPT volumes are provided in acquisition space or in atlas space. It is possible to extract data-specific links and metadata. From the collection of tract-tracing data connectivity matrices have been extracted, providing a coarse meso-scale connectome. Initial explorations indicate that the accuracy of these data is variable. A fundamental remaining question to resolve is whether it will be sufficient to index metadata and data links in the KG or whether image collections should be downloaded to HBP systems. The answer will depend on the functionality and design of tools for querying, extracting and visualising data from the KG (WP5.5). This will be addressed in collaboration with WP5.1 and WP5.5.

3.7.3.2 Component 39-2 Metadata and links to data shared via the mouse connectome project

Description of Component: Strategic data shared by the Mouse Connectome Project will be selected for ingestion into the KG in collaboration with researchers in SP4 and SP6. Required metadata will be entered using tools and forms provided in T5.4.1, and spatial metadata will be added using established tools (T5.4.2) and procedures (T5.2.2), curated, and released to the KG (http://www.mouseconnectome.org).

Progress on Component: Ongoing. Two repositories holding mouse brain tract-tracing data have been evaluated to determine whether their content can be made discoverable through the HBP KG:

1) The Mouse Connectome Project hosted at the University of Southern California (www.mouseconnectome.org) provides access to large bodies of mouse brain tract-tracing data (228 brains, 471 tracer injection sites) and associated metadata. Image data

are accessible through web-microscopy viewer tools showing tracing data together with normal cytoarchitecture and largely matching atlas diagrams from the Allen mouse brain reference atlas.

2) The Mouse Brain Architecture Project (http://mouse.brainarchitecture.org) hosted at the Cold Spring Harbor Laboratory provides access to 1,200 data sets holding mouse brain tract-tracing data with associated metadata, and >200 series of histological images showing brain-wide cell-type distributions. Microscopic images have been acquired as 3D volumes using serial two-photon microscopy, and can be inspected through an interactive image viewer system providing filters for regions of interest or specific cell types. Images are provided together with neighbouring Nissl stained sections and closest matching section diagrams from the Allen Mouse brain reference atlas.

Both resources provide access to a large amount of high-quality tract-tracing data mapped to the Allen Mouse brain atlas, and shared through interactive viewer tools. The spatial registration between section images and atlas plates is not perfect, but provides a useful starting point for analyses. Extraction of connectivity descriptions or matrices will require careful assessment of tracer injection sites and validation of the location of labelled axons. So far, detailed analyses of selected neural networks have been reported in the literature. Protocols for downloading raw image data are not provided. A next step is to contact database managers and clarify whether experimental and spatial metadata from these image collections can be indexed in the KG with specific links to original images or if image collections can be imported into HBP storage using SP5 workflows.

3.7.3.3 Component 40-3 Registration of rat brain hippocampal connectivity descriptions from temporal lobe database

Description of Component: Strategic data about the mouse shared via the Temporal Lobe repository (http://www.temporal-lobe.com) will be selected for ingestion into the KG in collaboration with researchers in SP4 and SP6. Required metadata will be entered using tools and forms provided in T5.4.1, and spatial metadata will be added using established tools (T5.4.2) and procedures (T5.2.2), curated, and released to the KG.

Progress on Component: Ongoing. The Temporal Lobe database provides an overview of all known anatomical projections of the rat hippocampal and parahippocampal regions, accumulated from 990 publications. References have been reviewed by experts in the field, and results have been organised in an interactive connectivity matrix. Original data are not directly accessible through the database. It is concluded that the Temporal Lobe database can be ingested in the HBP KG by relating semantic spatial metadata to the HBP ontology. The next step is to discuss implementation of this indexation with the HBP Tier 1 curation team.

3.7.3.4 Component 41-4 Registration of rat brain connectivity descriptions from the BAMS database

Description of Component: Strategic data about rat brain connectivity shared via the Brain Architecture Management System (https://bams1.org/)will be selected for ingestion into the KG in collaboration with researchers in SP4 and SP6. Required metadata will be entered using tools and forms provided in T5.4.1, and spatial metadata will be added using established tools (T5.4.2) and procedures (T5.2.2), curated, and released to the KG.

Progress on Component: Ongoing. The Brain Architecture Management System (BAMS) is no longer publicly available. Following several rounds of clarification, the content of the BAMS connectivity database content has been made accessible through KnowledgeSpace (WP5.8.3). The UIO team is now working with the KnowledgeSpace team to explore and evaluate the content of the BAMS connectivity database, in order to express key elements of the content through KnowledgeSpace searches. This in turn will allow content to be searchable through HBP KG.

3.8 T5.2.4 – Strategic Mining of Data Anchored to Rodent Atlases

3.8.1 *Key Personnel*

Task Leader: Jan BJAALIE (UIO)

3.8.2 *SGA1 DoA Goals*

Validate complete pipelines going from data ingested in the KG to interpretations of data and use of data for modelling and simulation.

This Task will focus on clarifying how different experimental data should be processed and analysed in order to yield quantitative data that are useful for subsequent predictive and modelling purposes. The data-mining use-cases will result in a first deliverance of quantitative data to data modellers from the HBP pipeline and valuable user feedback to tools developers and data curation teams.

3.8.3 *Component Progress*

3.8.3.1 Component 42-1 Region-wise key characteristics of neuronal morphologies

Description of Component: Strategic experimental data describing neuronal morphologies will be harvested from the pool of data accumulated in the KG. SP4 and SP6 researchers will be consulted about desired data specifications. Quantitative and semi-quantitative analyses will be conducted per brain region of interest to yield numbers (annotated with spatial metadata) that can be fed into analyses conducted by SP4 and SP6.

Progress on Component: Work on this Component will commence in M13. The starting point will be interactions with computational modellers in SP4 and SP6 to specify the type and format of quantitative data that can be useful for modelling purposes. Until the KG becomes operative, initial work will be conducted using data collected through efforts in T5.2 and T5.3. Analyses will be performed using a range of tools, including the Ilastik toolkit (WP5.6).

3.8.3.2 Component 43-2 Region-wise key characteristics of cellular elements

Description of Component: Strategic experimental data describing key cellular properties (densities, spatial distributions within subregions and other properties of different cell types) will be harvested from the pool of data accumulated in the KG. SP4 and SP6 researchers will be consulted about desired data specifications. Quantitative and semi-quantitative analyses will be conducted per brain region of interest to yield numbers (annotated with spatial metadata) that can be fed into analyses conducted by SP4 and SP6.

Progress on Component: Work on this Component will commence in M13. The starting point will be interactions with computational modellers in SP4 and SP6 to specify the type and format of quantitative data that can be useful for modelling purposes. Until the KG becomes operative, initial work will be conducted using data collected through efforts in T5.2 and T5.3. Analyses will be performed using a range of tools, including the Ilastik toolkit (WP5.6).

3.8.3.3 Component 44-3 Region-wise key characteristics of subcellular elements

Description of Component: Strategic experimental data describing key sub-cellular properties (densities, distributions of receptors, ion channels, synapses et cetera) will be harvested from the pool of data accumulated in the KG. SP4 and SP6 researchers will be consulted about desired data specifications. Quantitative and semi-quantitative analyses will be conducted per brain region of interest to yield numbers (annotated with spatial metadata) that can be fed into analyses conducted by SP4 and SP6.

Progress on Component: Work on this Component will commence in M13. The starting point will be interactions with computational modellers in SP4 and SP6 to specify the type and format of quantitative data that can be useful for modelling purposes. Until the KG becomes operative, initial work will be conducted using data collected through efforts in T5.2 and T5.3. Analyses will be performed using a range of tools, including the Ilastik toolkit (WP5.6).

3.8.3.4 Component 45-4 Region-wise key characteristics of connectivity

Description of Component: Strategic experimental data describing long-range neural connections will be harvested from the pool of data accumulated in the KG. SP4 and SP6 researchers will be consulted about desired data specifications. Quantitative and semi-quantitative analyses will be conducted per brain region of interest to yield numbers (annotated with spatial metadata) that can be fed into analyses conducted by SP4 and SP6.

Progress on Component: Work on this Component will commence in M13. The starting point will be interactions with computational modellers in SP4 and SP6 to specify the type and format of quantitative data that can be useful for modelling purposes. Until the KG becomes operative, initial work will be conducted using data collected through efforts in T5.2 and T5.3. Analyses will be performed using a range of tools, including the Ilastik toolkit (WP5.6).

3.9 T5.2.5 – Prediction of Cellular, Synaptic and Connectomic Composition, Distributions and Properties of the Rodent Brain

3.9.1 *Key Personnel*

Task Leader: Paul TIESINGA (SKU)

Other Researcher: Rembrandt BAKKER (SKU/JUELICH)

3.9.2 *SGA1 DoA Goals*

Develop an algorithm that can accurately predict the mesoconnectome based on publicly available gene expression data; provide a proof-of-principle implementation of this algorithm in MATLAB and evaluate the effectiveness of this algorithm.

In addition, we aim to reach MS5.2.4 during the reporting period, which is comprised of a prioritised roadmap for prediction targets.

3.9.3 *Component Progress*

3.9.3.1 Component 46-1 Synapse Generator Pipeline

Completed in the RUP.

3.9.3.2 Component 47-2 The Connectomic Composition Predictor

Description of Component: This Task will develop an algorithm to predict cell-type specific long-range projections, exploring and evaluating the performance of machine-learning techniques such as random forests, logistic regression, and Bayesian techniques to integrate the underlying multimodal and multilevel data. Pilot implementations will be provided in Matlab or R for ease of exploration by the recipient Use Cases. More robust implementations of the most effective algorithms will be in Python (and in C for the most computation intensive components) as part of follow-up work in SGA2. The input data derives from reconstructions of single cell axonal projections of thalamocortical neurons; the Allen mouse gene expression atlas together with single cell data obtained with RNA-seq and the Allen mouse mesoconnectome. The feasibility of including diffusion tensor imaging (DTI) data could also be explored. Deliverables associated with this product are a report on the algorithm and a review of the quality of the different underlying data sources, as well as a pilot implementation that will be delivered to the NIP as a downloadable Matlab toolbox.

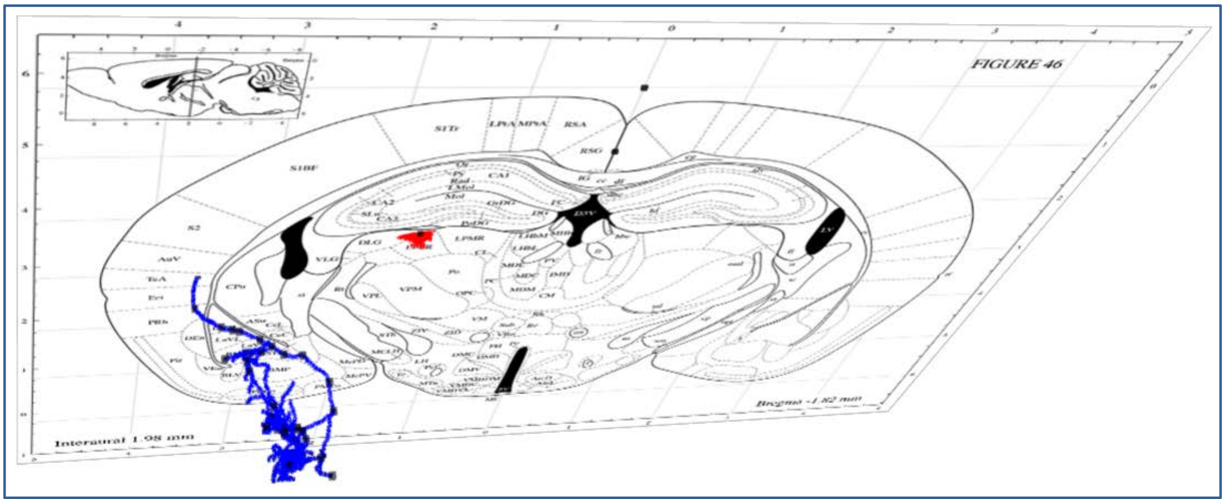

Figure 9: Projection neuron registered in the Paxinos and Franklin mouse atlas,[3] with dendrites in red and axons in blue. Screenshot from the HBP morphology viewer.

Further transformations based on the Allen institute mouse atlas will enable the embedding of the axonal reconstructions from the CLASCÁ laboratory into the NIP. A write-up describing the morphology viewer, its uses and the underlying data standard is at the time of writing in progress to ensure full exposure to prospective users.

To provide input to the Milestone document, a review[4] of the state-of-the-art in connectomic experimental techniques assessed their usefulness for the HBP consortium. We further wrote the Milestone document and are circulating it for comments by the involved partners. Briefly, we advise, given the available funding and publicly available data, to focus on improving the mesoconnectome using both missing data techniques, as well as gene expression data at the whole brain and single neuron level. The additional prediction target of the connectivity at cellular resolution would be most helped with new techniques that explicitly label the synapses and have the capability to feed into algorithms that can automatically reconstruct a large number of morphologies.

Ji *et al*., 2014, [5] claim that the mesoconnectome[6] can be predicted to 91% accuracy based on whole brain gene expression patterns[7] of only 25 genes. As part of the undergraduate internship of Roeland WIERSEMA we explored whether, in our hands, random forests, an ensemble technique, can also provide this accuracy. We were unable to achieve this high level of performance, which we believe is due to the approach of Ji *et al*., 2014,[5] to threshold the connectivity matrix for each sending area so that 50% of all possible connections are realised. Our aim is to choose a similar threshold for each sending area as this is the more principled procedure. We are at the time of writing recruiting a PhD student who can explore this matter further and implement a pilot tool in MATLAB. We have given the candidate this issue as test problem and are evaluating his performance to determine whether we can hire him for this task. The access to the gene expression data from the Allen institute will be provided via the NIP as part of task T5.2.4. While waiting until this capability becomes available, we have developed Python scripts that load these data directly from the Allen Brain Atlas. We assume that there is structure in the mesoconnectome. This structure can be used to improve the gene expression based prediction by correcting missing connections. To evaluate the power of this approach we have together with Max HINNE and Marcel van GERVEN at SKU developed a prediction method based on latent variables.[8]

Links:

Morphology viewer:

HBP Software Catalog:

https://collab.humanbrainproject.eu/#/collab/19/nav/2108?state=software,morphology-viewer

In the Collaboratory: https://collab.humanbrainproject.eu/#/collab/47/nav/4835

Direct URL: https://neuroinformatics.nl/HBP/morphology-miner/

Allen institute gene expression and connectivity data download scripts:

HBP Software Catalog:

https://collab.humanbrainproject.eu/#/collab/19/nav/2108?state=software,aba_mouse_connectivity_expression_download

Direct URL: https://neuroinformatics.nl/HBP/ABA_mouse/

4. WP5.3 Multi-level Atlas of the Human Brain

See Figure 10.

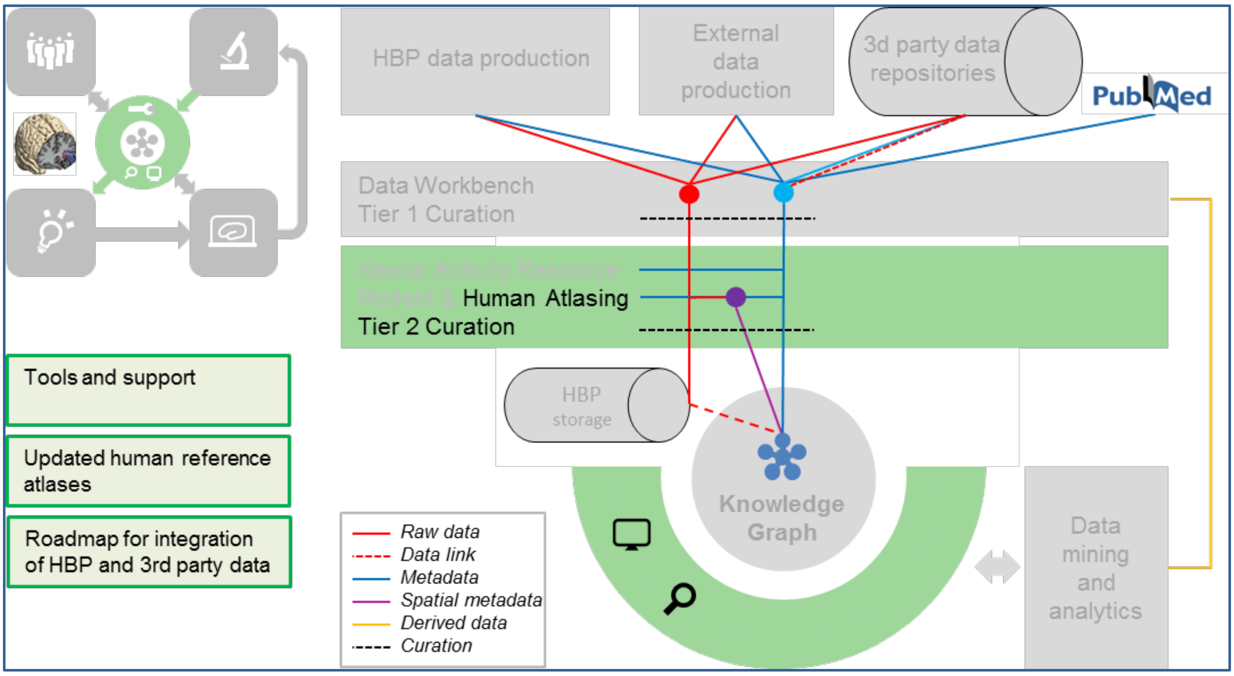

Figure 10: Schematic overview of contributions from WP5.3, highlighted in green.

4.1 Key Personnel

WP Leader: Timo DICKSCHEID (JUELICH)

Task Leaders: Timo DICKSCHEID (JUELICH)

Jean-François MANGIN (CEA)

Pedro LARRAÑAGA (UPM)

4.2 Work Package Leader's Overview

Despite delayed recruitment due to the changes in the SP5 work plan, JUELICH has worked on collecting human datasets delivered in the RUP from the data providers in SP2 and SP3, moving the data to sustainable storage in the SP7 computing centres, and starting to structure the metadata of these datasets. A first integration strategy has been defined for data from the Human Connectome Project (HCP), pushed by a fruitful Use Case with modellers in SP4. For T5.3.3, both JUELICH and UIO recruited prior to work plan approval, and fostered an early collaboration to kick-off the planned co-developments. T5.3.3 personnel in JUELICH was partly allocated for evaluating software solutions for T5.4.3 from M1-M12 to compensate the recruitment delays.

The data curation work in T5.3.1 is affected by late recruitment due to the delayed approval of the SP5 work plan, and the need to setup a new team. New personnel took up work at M13. Much of the work in WP5.3 has adjusted to the fact that the KG and FENIX storage infrastructure are not yet operational. Temporary solutions are used for storing the data in a fashion that will allow easy migration to the final systems, and connecting interactive tools to these data.

4.3 Priorities for the remainder of the phase

The priority is to get the new curation teams ready to assist with efficient integration of data into the atlas. Intensive coordination with WP5.1, WP5.2, and WP5.4 will take place, to agree on the metadata schemes, evaluate the feasibility of existing atlas tools for human data, and setup interim solutions for infrastructure components that are not yet operational. Temporary solutions must allow smooth transition to the final systems. A physical meeting of WP5.1 - WP5.4 is scheduled for 23 May in Geneva. Another focus in M12-M24 is to highlight the multimodal aspect of the atlas by prioritising the curation of the JuBrain cytoarchitectonic atlas as a key reference atlas, together with a selection of multimodal qualitative and quantitative datasets like fibre bundles and receptor distributions. Curators work closely with frontend developers in 5.4.3 on these priority datasets.

4.4 Milestones

See Table 3.

Table 3: Milestones for WP5.3 — Multi-level Atlas of the Human Brain

MS No.	Milestone Name	Lead Partner	Task(s) involved	Month Due	Month of Completion	Comments
MS5.3.1	List of KPIs for internal and external user adoption, of Products/services offered by WP5.3	20 (JUELICH)	T5.3.1-4	M10	M10	Revised and finalised by SP leader. Approved by Task/WP leaders. Minor adjustments of KPIs for internal SP use will be considered based on accumulated experience during the project period. A pre-verification version of this Milestone was delivered to the European Commission together with the D5.6.1 Resubmission in January 2017. The full list of KPIs is available for the HBP SGA1 consortium through the Milestone Report Repository and can also be found in Appendix 1. In this report, we include the M18 report on KPIs, with comments on adjustments of targets for SGA1
MS5.3.6	SGA1 and SGA2 roadmap for integration of HBP and third party human data into the human atlases. Validated by the CDPs and SP6	20 (JUELICH)	T5.3.1 T5.3.2	M12	M12	This roadmap is critical to planning of SGA2 efforts, and input from partners outside SP5 will be required. Several external data repositories have been evaluated, and a limited number of available example data sets downloaded and anchored to the reference atlas space. The roadmap has been commented on and approved by WP and Task leaders. A full report on this Milestone is available for the HBP SGA1 consortium through the Milestone Report Repository. https://emdesk.humanbrainproject.eu/shared/59a6619198092-fa7d99d6dff94d3d9f7e95371cab4cf3. At the time of submission of D5.8.2, milestones were uploaded in an HBP repository. Unfortunately, reviewers did not have access to this repository due to technical limitations

4.5 T5.3.1 - Curation of Human Atlas Data

4.5.1 *Key Personnel*

Task Leader: Timo DICKSCHEID (JUELICH)

4.5.2 *SGA1 DoA Goals*

The goal of this Task is to collect the relevant human datasets produced and delivered in SP2 (and partly SP3) from the data providers, and make it accessible through the NIP. This action usually comprises transfer of the data from the labs to sustainable HBP storage, curation of metadata into the NIPs metadata database, and spatial or semantic alignment to existing reference atlases. While the ultimate goal is to enable all of these steps through HBP services as they evolve (e.g. FENIX infrastructure services, curation services from 5.1, and atlasing tools from 5.4), T5.3.1 is also involved in co-designing these tools during SGA1.

4.5.3 *Component Progress*

4.5.3.1 Product 48-1 Whole Human Brain Cytoarchitectonic and Maximum Probability Maps registered and curated

Description of Component: The JuBrain cytoarchitectonic and maximum probability maps from the RUP will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the KG).

CDP to which Component contributes: This component contributes to CDP3-P3: Initial set of linked template datasets with labelled parcellation.

Progress on Component: The dataset represents a reference atlas, and comprises the MNI Colin 27 MRI volumes, individual probabilistic maps of roughly 100 different brain regions, as well as on maximum probability map representing a parcellation for all regions. All data are provided as Neuroimaging Informatics Technology Initiative (NIfTI) files. Curation and mapping of the metadata to NIP metadata schemes is in progress at the time of writing. This reference atlas has been integrated both into the RUP multi-resolution atlas viewer, and a preliminary release of the new 3D atlas viewer built in T5.4.3 (Figure 11).

Links:

- http://jubrain.fz-juelich.de
- http://www.fz-juelich.de/inm/inm-1/DE/Forschung/_docs/SPMAnatomyToolbox/SPMAnatomyToolbox_node.html

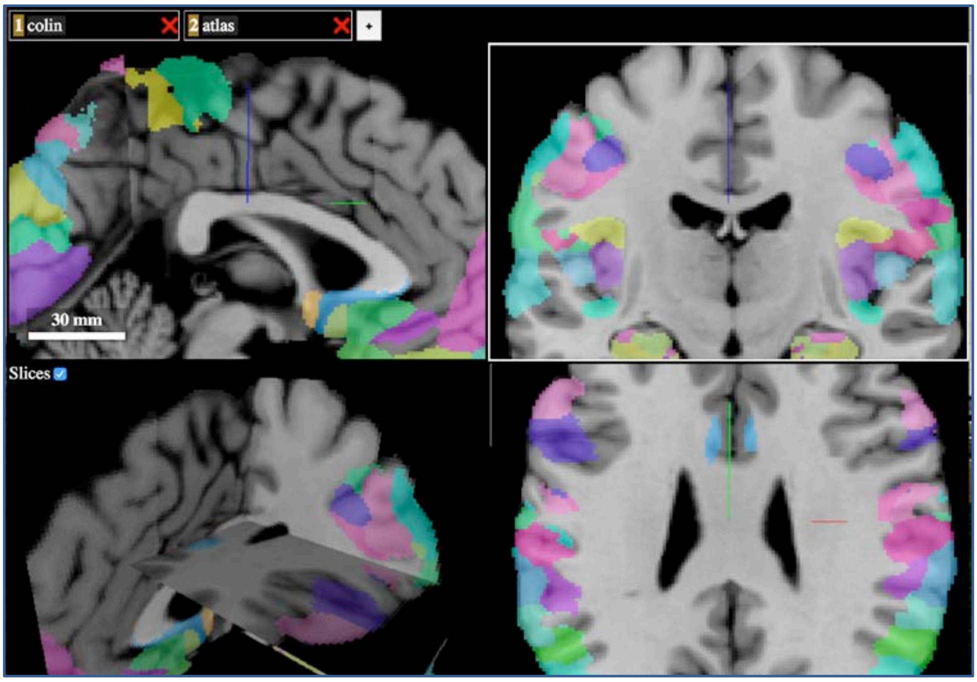

Figure 11: JuBrain cytoarchitectonic atlas integrated into a preliminary release of the new 3D atlas viewer (accessible here).

The parcellation refers to the maximum probability map of the JuBrain atlas.

4.5.3.2 Component 49-2 Human V1 laminar profiles from RUP registered and curated

Description of Component: Human V1 laminar profile of Figure Ground Segmentation (as produced by T2.1.8) will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the KG).

Progress on Component: The data has been collected from the data provider, and migrated to the GPFS file system in JUELICH that is part of the upcoming FENIX infrastructure. Curation of metadata has not yet started.

4.5.3.3 Component 50-3 BigBrain Release 2015 registered and curated

Description of Component: The 2015 release of the BigBrain will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the KG).

CDP to which Component contributes: This component contributes to CDP3-P3: Initial set of linked template datasets with labelled parcellation.

Progress on Component: The dataset represents a reference atlas, and comprises the histological volumetric data at different resolutions, a grey and white matter segmentation, and native mappings of the auditory cortex in a subset of coronal sections. The data has been integrated into the RUP multi-resolution atlas, and a preliminary release of the new 3D atlas viewer developed in T5.4.3 (Figure 12). The grey and white matter segmentation has been converted to Scalable Vector Graphics (SVG) format for display in the web viewer. Curation and mapping of metadata to the NIP metadata schemes are in progress at the time of writing.

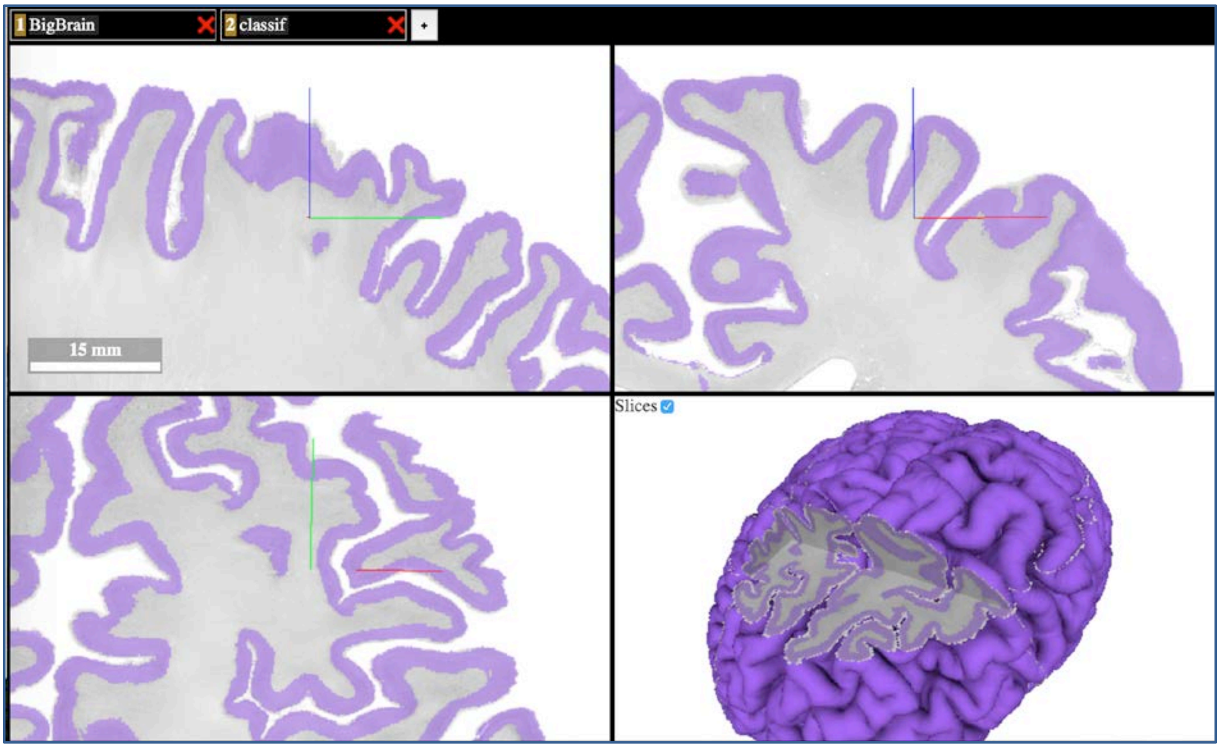

Figure 12: BigBrain 2015 release with grey and white matter segmentation integrated into a preliminary version of the new 3D atlas viewer build by T5.4.3.

Links:

- https://www.humanbrainproject.eu/en/explore-the-brain/atlases/
- https://bigbrain.loris.ca/main.php

4.5.3.4 Component 51-4 Infant atlas and major tracts in infant brains registered and curated

Description of Component: The RUP human dataset "Infant atlas and major tracts in infant brains" will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the KG).

CDP to which Component contributes: This component contributes to CDP3-P3: Initial set of linked template datasets with labelled parcellation.

Progress on Component: The data has been collected from the data provider, and migrated to the GPFS file system in JUELICH that is part of the upcoming FENIX infrastructure. The dataset represents a reference atlas and has been integrated into the RUP multi-resolution atlas viewer. Curation of metadata has not yet been started due to the delay in recruitment and pending implementation of the KG curation interfaces.

4.5.3.5 Component 52-5 Quantitative human receptor data in selected areas registered and curated

Description of Component: Quantitative human receptor data, as delivered from SP2 after the RUP, will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the KG). The same holds for the extended data to be delivered in M12 by MS MS2.2.1.

CDP to which Component contributes: This component contributes directly to CDP3-P4: Initial qualitative and quantitative datasets.

Progress on Component: The data has been collected from the data provider, and migrated to the GPFS file system in JUELICH that is part of the upcoming FENIX infrastructure. Example elements of this datasets have been used to demonstrate semantic linking in an atlas viewer by making them accessible through JuBrain region selection in the Papaya web-based viewer (Figure 13). The integration is realised via formalisation of the semantic links in csv and json files hosted by the Papaya web service. Curation of metadata has not at the time of writing started due to the delay in recruitment and pending implementation of the KG curation interfaces.

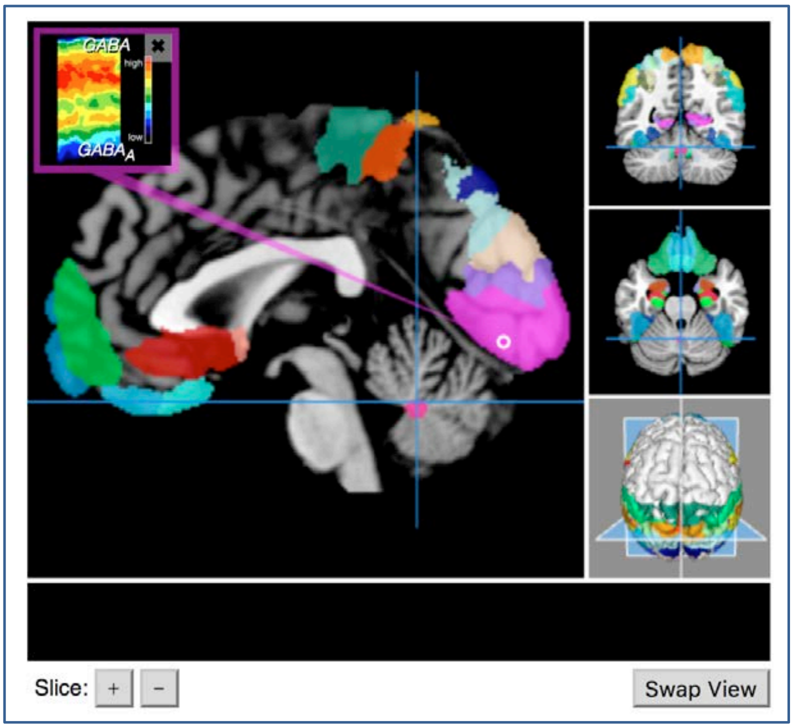

Figure 13: Receptor distribution plots semantically linked to the JuBrain atlas, demonstrated visually by the Papaya web-based viewer.

4.5.3.6 Component 53-6 Morphologies of selected human neurons registered and curated

Description of Component: Morphologies of selected neurons, as delivered from SP2 after the RUP, will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the KG).

Progress on Component: The data has been collected from the data provider, and migrated to the GPFS file system in JUELICH that is part of the upcoming FENIX infrastructure. Curation of metadata has not at the time of writing started due to the delay in recruitment and pending implementation of the KG curation interfaces.

4.5.3.7 Component 54-7 Whole brain connectivity atlas registered and curated

Note: The Component description has been updated. The Component description in the D5.6.1 Resubmission was erroneously only a copy of that for Component 53-6.

Description of Component: An atlas of long and short white matter structural connectivity, as delivered from SP2 after the RUP, will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the KG).

CDP to which Component contributes: This Component contributes to CDP3-P3: Initial set of linked template datasets with labelled parcellation

Progress on Component: The data has been collected from the data provider, and migrated to the GPFS file system in JUELICH that is part of the upcoming FENIX infrastructure. Curation and mapping of metadata to the NIP metadata schemes are in progress at the time of writing.

4.5.3.8 Component 55-8 Human Intracranial Database accessible in the NIP

Note: The Component description has been updated. The description in the D5.6.1 Resubmission erroneously referred to that for Component 54-7.

Description of Component: A collection of intracranial electroencephalogram (iEEG) data in 30 patients, each performing eight functional localisers, will be discoverable in the HBP atlas, and have their relevant metadata registered in the NIP metadata database.

CDP to which Component contributes: This Component contributes directly to CDP3-P4: Initial qualitative and quantitative datasets

Progress on Component: The data has been collected from the data provider, and migrated to the GPFS file system in JUELICH that is part of the upcoming FENIX infrastructure. Curation and mapping of metadata to the NIP metadata schemes are in progress at the time of writing.

4.5.3.9 Component 56-9 Wistar rat brain fibre orientation model registered and curated

Description of Component: The Wistar rat brain fibre orientation model delivered from SP2 after the RUP, will be discoverable in the HBP atlas, have a spatial correspondence to at least one of the accepted template spaces, and all relevant metadata registered in the NIP metadata database (i.e. what is today referred to as the KG).

Progress on Component: The fibre orientation model has been 3D reconstructed and registered into the Waxholm space in JUELICH. The location of the volume in the rat brain Waxholm space template has been verified as part of Tier 2 curation by T5.2.2 at UIO. Metadata registration into the KG will be done by T5.3.1, once the KG interfaces are available.

4.6 T5.3.2 Alignment and Import of Strategic Data from External Repositories

4.6.1 *Key Personnel*

Task Leader: Jeff MANGIN (CEA)

Other Researcher: Sandrine LEFRANCH (CEA)

4.6.2 *SGA1 DoA Goals*

Make publicly available datasets discoverable within the Atlas and seamlessly accessible to users through the atlas interfaces, as well as to HPC workflows that run on HBP infrastructure through NIP service APIs, as developed in WP5.5.

This Task will either create interfaces to important external repositories and atlases, or produce replicas on the federated HBP storage infrastructure. The initial focus will be to interface to the Allen Human Brain Atlas, using its well-defined API, and to replicate the HCP data (connectivity data for ~1,200 subjects). The reason for replicating the latter is that T5.4.4 is going to implement HPC workflows for applying established neuroimaging tools to these data. This is not feasible if the data are accessed from a remote site.

4.6.3 *Component Progress*

4.6.3.1 Component 57-1 Software for external datasets import/integration

Description of Component:

a) Integration via APIs. Data are not duplicated, but rather accessed directly on the Allen/HCP servers. Web-based APIs will be developed to on-the-fly transform metadata and data from original representation to a representation suitable for HBP.

b) Dataset replication. Data are replicated on HBP repositories (exploiting FENIX), and transformed to HBP standard formats for data/metadata. The conversion in this case is offline.

Progress on Component: Up to now we have concentrated on the processing of the data required for integration. Software for integration in the NIP will be developed over the coming months.

4.6.3.2 Component 58-2 Human Connectome Project data searchable in NeuroInformatics Platform (data)

Description of Component: Data from the HCP are fully accessible and searchable from NIP. Metadata comply with HBP standards. The user can navigate the data using software developed in T5.5.2, and can use these data for analysis or simulation. Here, the data are replicated on HBP repositories (exploiting FENIX), and transformed to HBP standard formats for data or metadata. The conversion in this case is offline.

CDP to which Component contributes: This Component contributes to CDP3 (the human brain atlas) via the connectivity matrix exportation Use Case.

Progress on Component: The latest release of the HCP dataset (March 2017) has been downloaded in CEA and in JUELICH to be integrated in FENIX from JUELICH'S HPC. One postdoc was sent to the 2016 HCP summer school to learn the optimal use of the functional magnetic resonance imaging of the brain software library (FSL) distributed to preprocess the data. The different analysis pipelines that will be used to integrate the dataset in the NIP have been designed and tuned to the outstanding nature of the data. They will provide a variety of connectivity matrices, based on standard and HBP specific parcellations. Users will be able to evaluate the influence of pipeline parameters, either using a wide set of existing pipeline runs archived in FENIX or requiring new processing performed by SP7 facilities.

4.6.3.3 Component 59-3 Allen Human Brain Atlas data searchable in NeuroInformatics Platform (data)

Description of Component: Data from Allen Human Brain Atlas are fully accessible and searchable from the NIP. Metadata comply with HBP standards. The user can navigate the data using software developed in T5.5.2, and can use these data for analysis or simulation.

CDP to which Component contributes: This Component contributes to CDP3 (the human brain atlas).

Progress on Component: The strategy to embed the Allen Human Brain Atlas consists in computing gene expressions for the different ROI of the HBP atlas. A first level integration based on the Jubrain cytoarchitectonic cytoarchitecture atlas is at the time of writing in progress in JUELICH, in cooperation with T2.6.5.

4.7 T5.3.3 – Cross-scale Interactive Spatial Alignment Tools for Partial Volumes

4.7.1 *Key Personnel*

Task Leader: Timo DICKSCHEID (JUELICH)

Other Researcher: Jan BJAALIE (UIO)

4.7.2 *SGA1 DoA Goals*

To implement an initial web-based solution to manually align a high-resolution partial volume to a reference atlas volume.

The workflow will be such that: 1) the partial volume file is uploaded to the server; 2) the user is able to enter 3D corresponding landmarks between the two volumes, interacting with two synchronised tri-planar views of the two volumes; and 3) the landmarks are used to compute and apply a 3D affine transformation that aligns the partial volume with the reference dataset. The workflow will allow iterative refinement of the landmarks, by providing the possibility to cycle through steps 2 and 3. Once implemented, the tool facilitates spatial anchoring of high-resolution data to the atlases, and hence supports future Tier 2 curation tasks.

4.7.3 *Component Progress*

4.7.3.1 Component 60-1 Web-based multi-resolution three-planar viewer for large image volumes

Description of Component: This is a web-based Component that presents the user with three synchronised orthogonal planar views (coronal, axial, and sagittal) of an image volume, and a cross-cursor to pinpoint a particular point within this volume. A multi-resolution data streaming strategy is used, in order to allow for remote visualisation of reference volumes that do not fit into the client-side working memory. This product will either build on the big data viewer from T5.4.3 (Product 89-1), or apply the same 2D web rendering engine as Product 85-11. It will use the HBP image service (Product 101-1) as a back-end for serving the multi-resolution data. It will be used by Product 61-1, as well as Product 87-13, as a component of their user interface for entering the landmarks.

CDP to which Component contributes: This Component contributes to CDP3-P5: Alignment tools for incoming data.

Progress on Component: A user interface (layout) concept and pilot for the three-planar viewer has been successfully developed and shared between UIO and JUELICH. Test data for this Component has been preprocessed into multi-resolution format ("Loris", 278 Gvoxels, ended up as 200 Gbytes of pyramidal tiles) with *ad-hoc* infrastructure.

4.7.3.2 Component 61-1 Selection, management and navigation of many landmarks

Description of Component: This Component will implement a web-based HTML5 interface that allows interactive input and management of many 3D landmarks, i.e. explicit corresponding points in two synchronised views: a view of the incoming partial 3D volume that is being spatially anchored, and a view of the reference template volume. Either volume should be visualised through an image service which supports streaming of multi-resolution data, in order to allow for very large volumes. This Component is based on Product 60-1, which provides three orthogonal planar views into the incoming and template volumes. Development of the front-end will be shared with Product 87-13, which also needs an interface for entering of corresponding 3D landmarks in two volumes.

CDP to which Component contributes: This Component contributes directly to CDP3-P5: Alignment tools for incoming data.

Progress on Component: An initial implementation of a web-based frontend for the landmark adjustment has been implemented in Javascript (Figure 14). The code has been added to an HBP-internal GitHub repository that is shared between UIO and JUELICH for co-development. Integration of the three-planar viewer components (product 60-1) with this frontend is in progress at the time of writing (Figure 15). Practical co-development between the two groups – JUELICH and UIO – has thus been started, and is working successfully.

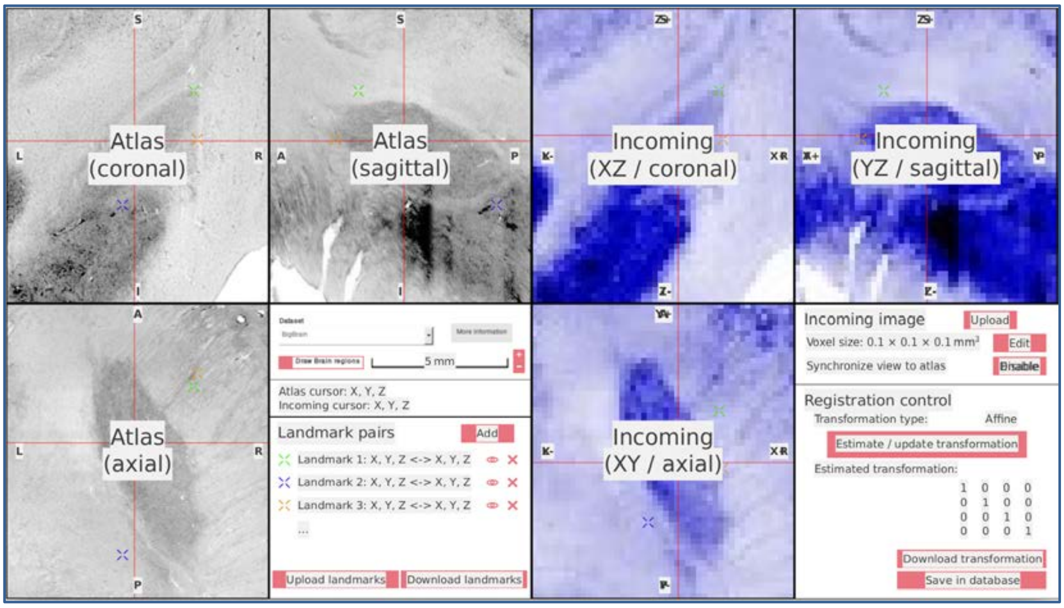

Figure 14: Prototype implementation of a JavaScript-based frontend for corresponding landmark adjustment between two tri-planar views of a source and target volume.

The image views are design elements from M12 used in the planning of the tool development.

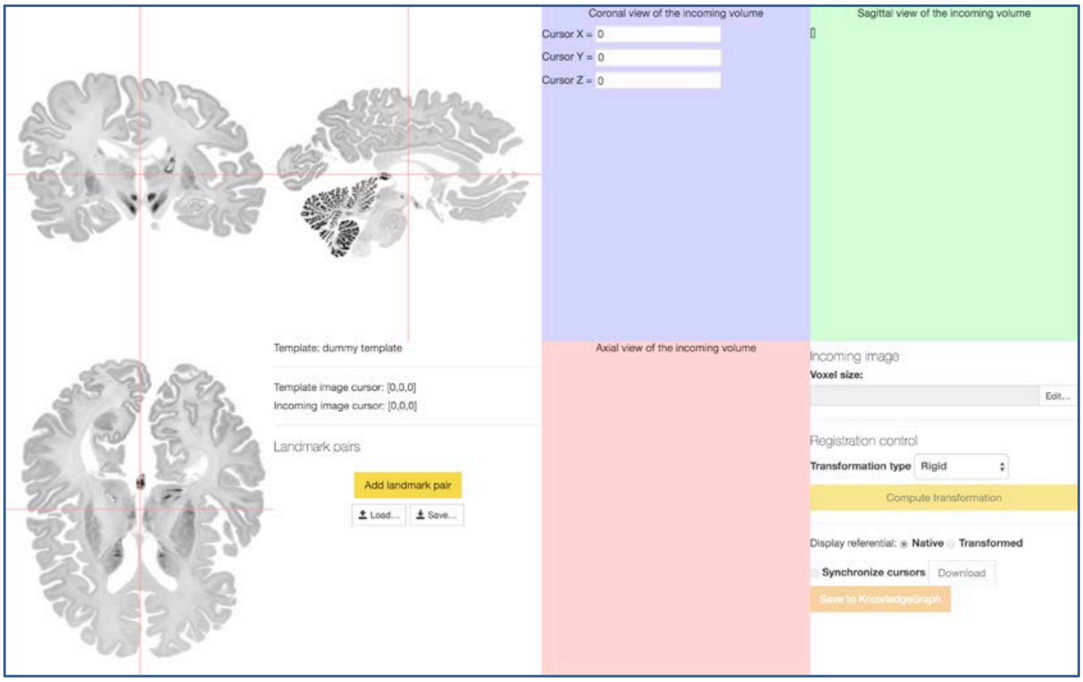

Figure 15: Early integration of the three-planar viewer (product 60-1, implemented at UIO) with the frontend for landmark adjustment (product 61-1, implemented at JUELICH).

High-resolution image data is streamed via an http service to the frontend.

4.7.3.3 Component 61-2 Affine transformation estimation from landmarks

Description of Component: This Component will implement a back-end that can derive an optimal 3D affine transformation from a set of explicit pointwise landmark correspondences, entered using the front-end developed in the previous Product 61-1. This is different from Product 87-31, which provides intensity-based registration for whole brain data from rodent. However, both tools will share code where possible, e.g. for actually applying the estimated transformations to the source image.

CDP to which Component contributes: This Component contributes directly to CDP3-P5: Alignment tools for incoming data.

Progress on Component: A REST API for communication between the front-end and back-end has been designed, which will allow transference of a list of corresponding landmarks, return spatial transformation parameters, and trigger re-sampling of the registered image. A technical review has been conducted for choosing a suitable Python framework to implement this API.

4.7.3.4 Component 61-3 Iterative workflow loop for landmark adjustment

Description of Component: This product will implement a controlled interactive loop, during which corresponding landmarks can be updated, added, or deleted in order to refine the registration. The effect on the registration made visible after each update, and the option is given to the user to re-slice the incoming partial volume in order to align its axes with the axes of the template volume.

CDP to which Component contributes: This Component contributes directly to CDP3-P5: Alignment tools for incoming data.

Progress on Component: As a preliminary step, an iterative workflow loop has been successfully prototyped as a script around the existing QuickNII tool (intended for use with rodent data) from the RUP as a mock-up solution to study the overall feasibility and user experience, and to arrive at a common understanding of the problem.

4.7.3.5 Component 61-4 Connection and interoperability NIP services

Description of Component: This Task will link the alignment tool to the KG and HBP data services, including a streaming image service (Product 101-1), so that reference templates from the HBP atlas can be naturally accessed.

CDP to which Component contributes: This Component contributes directly to CDP3-P5: Alignment tools for incoming data.

Progress on Component: Not yet started. Connection to NIP services will be the last step of the implementations in SGA1, and requires the availability of HBP image services (T5.5.4), complete curation of the respective reference atlases into the metadata database, and availability of a mature API to the metadata database.

4.8 T5.3.4 - Machine-Learning and Statistical Methods for Modelling Cellular and Subcellular Morphologies

4.8.1 *Key Personnel*

Task Leader: Pedro LARRAÑAGA (UPM)

4.8.2 *SGA1 DoA Goals*

Develop machine-learning models to identify morpho-electrical neuron types using unsupervised classification and multi-output regression, and model the geometry of human pyramidal cells.

This involves:

(a) Hierarchical basal arborisation models covering from its orientation to the detailed dendritic branching of each arbour. Sampling from the Bayesian network models will generate virtual human pyramidal cells.

(b) Repairing models of incomplete basal dendrites.

(c) Analysis of dendritic wiring optimality. Heuristic optimisation will test whether wiring economy is related to the way in which dendrites grow. A third more fine-grained goal is to provide a morphological classification of dendritic spines from a probabilistic mixture-based clustering approach that in addition will be able to simulate virtual spines from each class.

This Task will focus on cortex pyramidal human cells and will be done in collaboration with Idan SEGEV (T6.2.1) and Javier DEFELIPE (T1.2.1).

4.8.3 *Component Progress*

4.8.3.1 Component 62-1 3DSynapsesSA

Completed in RUP.

4.8.3.2 Component 63-1 3DSomaMS

Completed in RUP.

4.8.3.3 Component 64-3 3DPyrStructure

Completed in RUP.

4.8.3.4 Component 65-5 Single cell arborisation model

Description of Component: Hierarchical Bayesian model of the basal arborisation morphology built from 3D pyramidal neuron reconstructions.

Use Case: Basal arborisation single-cell hierarchical models.

Progress on Component: The first release on this Component is planned for M21; we have thus performed no work so far on this Component.

4.8.3.5 Component 66-5 Incomplete basal dendrites repairing model

Description of Component: Data-driven repairing model that detects cut-points in the basal arborisation and then repairs them using a growth model built from complete 3D neuron reconstructions.

Progress on Component: We have had no planned releases in M7-M12. We are currently computing the set of features needed to characterise basal dendrites.

4.8.3.6 Component 67-6 Spine morphology feature extractor

Description of Component: MATLAB module that given a 3D spine reconstruction computes a set of characteristic morphological measures that unequivocally determine the spine shape. See Figure 16.

Progress on Component: We have released the public version of the software. The software computes a set of morphological features for each dendritic spine and saves the measures in a csv file. It is registered in the Software Catalog (3DSpineMFE) and available at the following designated GitHub repository.

Links: https://github.com/ComputationalIntelligenceGroup/3DSpineMFE

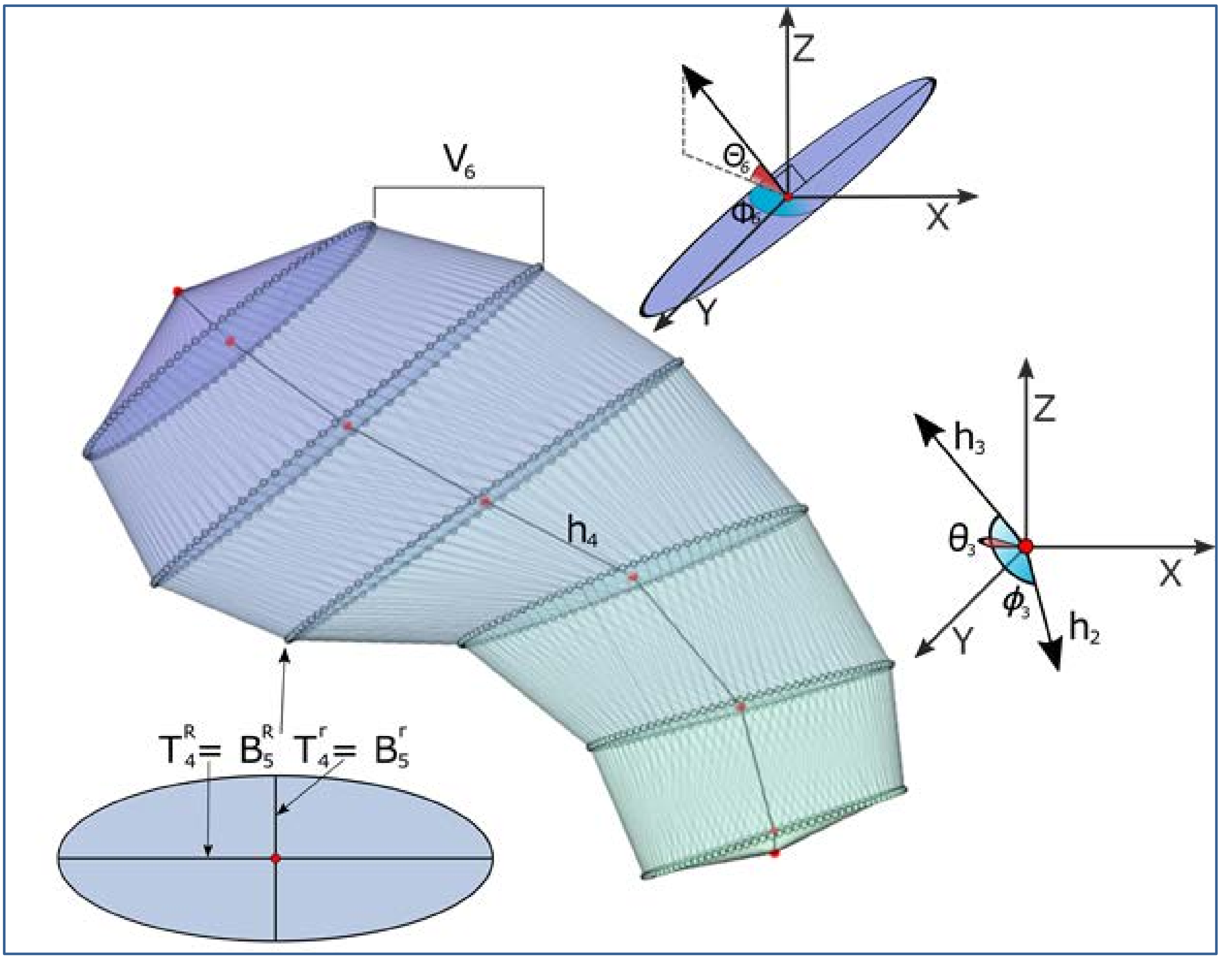

Figure 16: An ellipse is defined by its centroid, major axis ($T^R_{i-1} = B^R_i$) and minor axis ($T^r_{i-1} = B^r_i$).

These points are connected by vectors h_i whose length is $|h_i|$. From vectors h_i and h_{i-1}, θ_i and ϕ_i are obtained. Θ_i and Φ_i are the ellipse directions of the spine.

4.8.3.7 Component 68-7 Spine morphology clustering

Description of Component: Probabilistic clustering model that discovers groups of spines according to their morphological features.

Use Case: Dendritic spines morphological classification.

Progress on Component: We have had no planned releases in M7-M12. We are starting the modulisation and clustering tasks.

4.8.3.8 Component 69-8 Spine morphology simulator

Description of Component: R package that samples from a probabilistic model to create realistic 3D virtual spine representations that match morphologically to the real ones.

Progress on Component: We have had no planned releases in M7-M12. Output of 2.3.5 is needed to start this Component.

4.8.3.9 Component 70-9 Morpho-electrical neuron types component

Description of Component: Model that predicts unseen neurons electrophysiological variables from the set of morphological ones. The model has to be validated via regression performances. Apart from the prediction task, we want to learn models that describe relationships among variables, giving thoughtful insights in neurons behaviour.

Progress on Component: We have had no planned releases in M7-M12. We are collecting data from HBP partners (requested) and from publicly available sources (e.g. Allen institute Cell types database). We started the exploratory analysis of the available data, waiting to be able to access the data from the HBP partners.

5. WP5.4 Data and Atlas Curation Tools

See Figure 17.

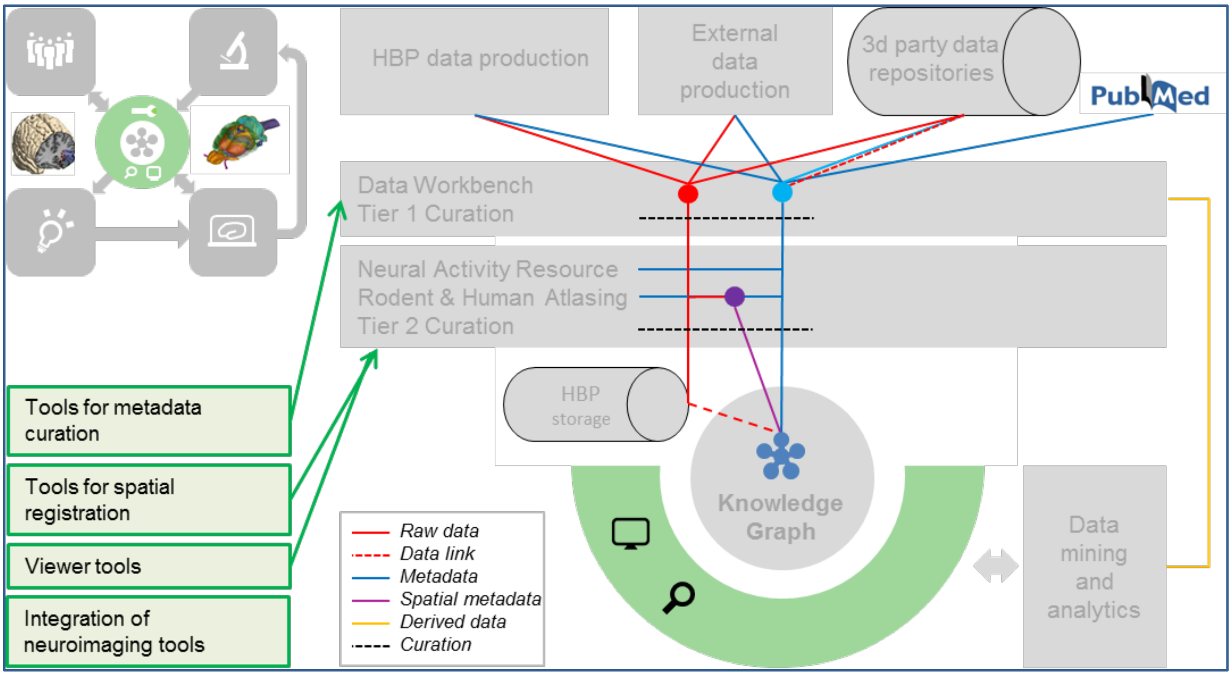

Figure 17: Schematic overview of contributions from WP5.4, highlighted in green.

5.1 Key Personnel

WP Leader: Jan BJAALIE (UIO)

Task Leaders: Martin TELEFONT (EPFL)

Jan BJAALIE (UIO)

Timo DICKSCHEID (JUELICH)

Simon EICKHOFF (JUELICH)

5.2 Work Package Leader's Overview

Progress has been made on the majority of tools to be developed under WP5.4. The QuickNII tool, critical for anchoring images to reference atlases, has been completed. A prototype for the 3D high-volumetric interactive atlas viewer has been developed. Based on the prototype, initial releases of static atlas viewers for human and rodent reference atlases have been finalised (e.g. http://bigbrain.humanbrainproject.org). Significant progress has been made on establishing preprocessing pipelines for raw neuroimage data on JUELICH's JURECA supercomputer.

Recruitment has now been finalised and all product developments have started.

Based on the current status of this WP, it is expected that tools under this WP will be delivered at the planned level by the end of SGA1, covering preprocessing of raw neuroimage data, anchoring of image data to reference atlases, and viewing of reference atlases and anchored data in atlas space. The QuickNII tool for anchoring of data to reference atlases, with new functionality for propagating spatial transformations across large series of 2D images, is heavily used in the workflow for building the multi-level atlas of the rodent brain (WP5.2).

5.3 Priorities for the remainder of the phase

In the remainder of the phase, priority will be given to complete all product developments.

5.4 Milestones

See Table 4.

Table 4: Milestones for WP5.4 — Data and Atlas Curation Tools

MS No.	Milestone Name	Lead Partner	Task(s) involved	Month Due	Month of Completion	Comments
MS5.4.1	List of KPIs for internal and external user adoption of Products/services offered by WP5.4	81 (UIO)	T5.4.1-4	M10	M10	Revised and finalised by SP leader. Approved by Task/WP leaders. Minor adjustments of KPIs for internal SP use will be considered based on accumulated experience during the project period. A pre-verification version of this Milestone was delivered to the European Commission together with the D5.6.1 Resubmission in January 2017. The full list of KPIs is available for the HBP SGA1 consortium through the Milestone Report Repository and can also be found in Appendix 1. In this report, we include the M18 report on KPIs, with comments on adjustments of targets for SGA1

5.5 T5.4.1 – Tools for Metadata Curation

5.5.1 *Key Personnel*

Task Leader: Martin TELEFONT (EPFL)

5.5.2 *SGA1 DoA Goals*

Develop, deploy and maintain software essential to data discoverability and accessibility across the HBP Platforms.

This Task will use open source code where appropriate, and the software will be released as open source from the first code commit. User Support based on these tools is provided in WP5.7. Input from WPs managing various categories of data (WP5.2, WP5.3, and WP5.7) will be used to develop unique metadata schemas for each data category. The following functionalities will be provided:

- A WebApp, and APIs, for Meta-Data Annotation of Data to be saved in KG;
- A WebApp, and APIs, for Curation of Parameters;
- A webclipper, for manual parameter extraction from web resources;
- A WebApp, and APIs, for Ontology Management and expansion;
- A WebApp, and APIs, for a Data-Standards Catalogue;
- A WebApp, to search and navigate HBP, and community, Data and Parameters. Users will be able to put a "Shopping Basket", to build data and model collections, which can then be shared, or exported for processing/display in the HBP Platforms, and standalone apps;
- A WebApp, and APIs, to easily deposit and access data and models contributed to the Neural Activity Resource (NAR) (developed with T5.7.2) as well as the underlying database software infrastructure.

5.5.3 *Component Progress*

5.5.3.1 Component 71-1 Registration in KnowledgeGraph

Description of Component: Datasets are annotated with high-level metadata and stored in a database.

Progress on Component: Planning efforts completed and mapping of the metadata collected under WP5.1 into the KG has started. See detailed explanation under WP5.1 and WP5.5.

5.5.3.2 Component 72-2 Data Workbench (API, WebApp, MetaData DB)

Description of Component: Data Workbench allows users to enrich their data artefacts with metadata, so as to make them discoverable via a search interface (API and web). While annotating the data users will contribute to the development of ontologies where concepts are missing, and feed a data-standard database.

Progress on Component: Initial versions have been tested with internal Tier 1 (WP5.1) and Tier 2 (WP5.2 and WP5.3) curation teams resulting in extensive change requests which are in the process of being implemented. Initial efforts in ontology alignment with KG teams has started.

5.5.3.3 Component 73-3 Parameter Workbench (API, WebApp, MetaData DB)

Description of Component: Parameter Workbench allows users to capture parameters and annotate them with metadata, using the same ontology as in the Data Workbench, so as to make them discoverable via a search interface (API and web). The Parameter Workbench will also allow users to discuss with other uses the likelihood that parameters reported in the literature and online databases can be taken at face value, or how they might need to be adjusted to reflect the current state of knowledge.

Progress on Component: Initial planning.

5.5.3.4 Component 74-4 Neural Activity Resource Development (API, WebApp, MetaData DB)

Description of Component: For integration and analysis of neural dynamics data [ECoG, functional MRI (fMRI), cellular, ensembles, model results, behaviour, etc.] Products produced in T5.1.1 and operationally supported by the work in T5.7.1. The development in T5.7.2 is the in-depth adaption of the resource together with the users.

Progress on Component: A user interface (layout) concept and pilot for the three-planar viewer has been successfully developed and shared between UIO and JUELICH. Test data for this component has been preprocessed into multi-resolution format ("Loris", 278 Gvoxels, ended up as 200 Gbytes of pyramidal tiles) with ad-hoc infrastructure.

5.6 T5.4.2 – Integrating 2D Atlas Viewers and Manual Spatial Registration Tools

5.6.1 *Key Personnel*

Task Leader: Jan BJAALIE (UIO)

5.6.2 *SGA1 DoA Goals*

This Task will include development of:

- a tool for anchoring of 2D experimental image data to 3D reference atlases, with functionality to increase the speed of the anchoring process through optimised user interfaces and propagation of anchoring information through large series of images;
- a tool for 2D to 2D non-linear warping, allowing the user to add a second step to the anchoring process, achieving a higher level of precision when required;
- a tool for simultaneous viewing of individual section images with overlay of the relevant atlas cut plane following anchoring to reference atlas, including functionality for reading spatial coordinates from the reference atlas and adding annotations in reference atlas coordinate space;
- a back-end tool for 3D to 3D non-linear warping allowing registration of whole brain 3D volumes to the reference atlases. The user will access this tool through the front-end software developed under T5.3.3.

Tools developed under this Task will be part of a pipeline starting with the curation of metadata required to perform the anchoring of images to reference atlases and ending with the retrieval of data for viewing, annotation and analysis. All tools will be adapted to the different data categories produced in SP1 and the data standards developed by the Ontology Definition Team (ODT) to be realised in T5.1.2 and supported in T5.1.1.

5.6.3 *Component Progress*

5.6.3.1 Component 75-1 AligNII, from RUP: online tool for anchoring of 2D experimental image data to 3D reference atlases

Completed in RUP.

5.6.3.2 Component 76-2 QuickNII, from RUP: standalone tool for anchoring of 2D experimental image data to 3D reference atlases

Completed in RUP.

5.6.3.3 Component 77-3 CutNII, from RUP: custom-angle slicer for 3D reference atlases

Completed in RUP.

5.6.3.4 Component 78-4 MeshGen, from RUP: Tool for generating smooth surface meshes from volumetric segmentation data

Completed in RUP.

5.6.3.5 Component 79-5 MeshView, from RUP: Online 3D surface and custom slice viewer for reference atlas mesh data

Completed in RUP.

5.6.3.6 Component 80-6 Mesh set for Waxholm Space rat brain atlas (from RUP)

Completed in RUP.

5.6.3.7 Component 81-7 Mesh set for Allen mouse brain reference atlas (from RUP)

Completed in RUP.

5.6.3.8 Component 82-8 Mesh set for Waxholm Space mouse brain reference atlas (from RUP)

Completed in RUP.

5.6.3.9 Component 83-9 QuickNII v 2.0: updated functionality and new procedures for propagation of anchoring information through large series of images

Description of Component: The next generation of the QuickNII software (v 2.0) will be optimised for faster anchoring of large series of 2D images to the reference atlases. The new version will build on the version from the RUP (product 76-2) and will include new functionality for propagating spatial transformations across series of sections following anchoring of selected images.

Progress on Component: Completed. Following several pilot projects and extensive testing, implementation of the propagation feature is complete. The feature will be extracted and made available as a standalone component for reuse.

Available in the Software Catalog:

Link: https://collab.humanbrainproject.eu/#/collab/19/nav/2108?state=software,quicknii

5.6.3.10 Component 84-10 Non-linear warping of anchored 2D images to a reference atlas

Description of Component: Anchoring (affine registration) of section images to a reference atlas is performed using QuickNII (product 83-9). This Component will provide a new functionality for enhancing the registration in the reference atlas by applying a 2D to 2D non-linear approach. These non-linear techniques will be based on image intensity features or user-defined landmarks features. The possibility of using a non-linear technique and the choice of the non-linear technique (intensity-based or landmark-based or both) will be dependent on the type of material and the contrast in the section images. The Component will be considered for inclusion in an extended version of QuickNII.

Progress on Component: Proof of concept has been developed based on *bUnwarpJ* tool from ImageJ allowing combination of landmarks or image intensity. This tool has been applied to two 2D histology datasets from RUP (Whole brain connectivity) anchored to the Waxholm space. Improvement of the localisation of the regions is observed for these two datasets.

5.6.3.11 Component 85-11 LocaliZoom: viewer for series of 2D images with reference atlas superimposed

Description of Component: Web-viewer tool for viewing of series of 2D images that have been anchored to reference atlases. The tool allows display of the relevant reference atlas cut planes superimposed on the images at a user-defined level of transparency. The tool will have additional functions for graphical and semantic annotation functionality and reading of spatial coordinates (Waxholm Space and Bregma coordinates) for points of interest in the images. Spatial coordinates can be exported to MeshView v 2.0 (product 86-12).

Progress on Component: Base functionality has been implemented (considered feature-complete) as a specialised component for handling histological sections anchored with QuickNII. LocaliZoom is integrated with the internal Tier 2 atlas curation system at UIO. Integration with HBP resources/systems is ahead.

5.6.3.12 Component 86-12 MeshView v2.0: updated functionality, viewing of annotations from LocaliZoom

Description of Component: The next generation of the MeshView web-viewer for interactive viewing of volumetric vector-based meshes from reference atlases and cutting of the reference atlas volumes in arbitrary, user-defined planes, providing customised atlas plates. The new version will provide functionality for viewing of annotations from LocaliZoom (product 85-11). MeshView v2.0 thus delivers results aggregated from series of 2D images, anchored to reference atlas using QuickNII and annotated in LocaliZoom.

Progress on Component: MeshView has been extended with the base functionality for displaying point clouds. It covers the Use Case of displaying annotations from LocaliZoom (which are points), and also works with point clouds from other data sources. Remaining tasks are as follows: cutting for point clouds (now they work best with transparency) and integration with other applications (point cloud data is transferred via copy-paste in the first version)

5.6.3.13 Component 87-13 Non-linear warping of whole brain 3D volumes to reference atlas

Description of Component: Tool for intensity-based non-linear registration of whole brain 3D volumes to the 3D reference atlases, aimed at managing data sets from rodent brain. For visualising and controlling the registration, the front-end developed as Product 60-1 and 60-2 of T5.3.3 for human brain data will be co-developed, integrated and used.

Progress on Component: Proof of concept of multi-level multi scale normalisation procedure for whole brain rodent volumes has been developed based on Advanced Normalisation Tools (ANTs). This tool has been applied to normalise one high resolution (40 µm) 3D MRI volume to the Waxholm MRI template. Three other datasets have been identified and will be processed in a similar fashion.

5.6.3.14 Component 88 -14 Transformation inverter

Description of Component: This software methodology will invert the transformations established by affine and non-linear spatial registration of customised 2D atlas slices to images (using Products 83-9 and 84-10) and apply these to transform 2D images to reference atlas space.

Progress on Component: Not started.

5.7 T5.4.3 – Development of 3D High-Volumetric Interactive Atlas Viewer

5.7.1 *Key Personnel*

Task Leader: Timo DICKSCHEID (JUELICH)

5.7.2 *SGA1 DoA Goals*

Set up web-based solution for exploring in 3D large volumetric data that does not fit into the client memory.

The data are assumed to be provided as a specifically prepared http-hosted pyramidal file format, or through the HBP image service being developed by 5.5.4. As a key Use Case and challenge, we aim for a fluent and intuitive web-based visualisation of the BigBrain dataset (20 micron isotropic), together with a classification of its grey and white matter.

5.7.3 *Component Progress*

5.7.3.1 Component 89-1 Web based big data viewer for navigating the BigBrain in three planes at different resolutions

Description of Component: Web-based viewer for high-resolution BigBrain data with capabilities for interactively panning and zooming the image data in three different planes.

CDP to which Component contributes: This Task directly contributes to CDP3-P2: 3D interactive big data viewer.

Progress on Component**:** Due to the late approval of the SP5 work plan, recruitment for this Task started late (M9-M11). A number of steps towards the development of the new viewer were described in our initial report, including an evaluation of a range of existing solutions for 3D visualisation of large datasets (see feature comparison table in Figure 18). The Neuroglancer solution was preferred, since it already provides compatibility with popular image service standards, promising performance, and good uptake by relevant communities. Since then, a first version of a web-based 3D atlas viewer for terabyte-sized brain volumes has been implemented. The viewer is based on Neuroglancer and can display volumetric data with parcellations of brain structures on top. It has been extended to provide an intuitive whole brain 3D overview with surface meshes and front octant removal (inspired by the Atelier3D software) as well as an API that allows to use it as an embedded component in interactive web applications. The extension is called NeHuBa (Neuroglancer for human brain atlas). Explicit releases are now available based on this software for the human atlases (JuBrain atlas and the BigBrain template) as well as for the rodent reference atlases. A development version of an interactive viewer which embeds NeHuBa, and allows templates and parcellations of brain regions to be selected, has been tested.

Links:

Release of the viewer has been deployed to HBP webservers and hosts initial rodent and human reference atlas datasets:

https://www.humanbrainproject.eu/en/explore-the-brain/atlases/

The Docker image of the Neuroglancer:
https://github.com/HumanBrainProject/neuroglancer-docker

3D Atlas Viewer:
https://collab.humanbrainproject.eu/#/collab/19/nav/2108?state=software,atlas-viewer-neuroglancer

Conversion scripts for Neuroglancer:
https://collab.humanbrainproject.eu/#/collab/19/nav/2108?state=software,neuroglancer-conversion-scripts

Feature	Importance of the feature for a 3D atlas viewer	Neuroglancer	Shadernavigator	openseadragon	Papaya
Ability to display cross-sectional views (obliques)	absolutely required	yes	yes	n/a	no
Ability to display images larger than client-side working memory	absolutely required	yes	yes	yes	no
Ability to display volumetric data	absolutely required	yes	yes	no	yes
Display of 3D surface meshes	absolutely required	yes	no	n/a	yes
display of indexed images (label images)	absolutely required	yes	?	yes	yes
Displays both voxel and surface data within the same space coordinate system	absolutely required	yes	no	n/a	yes
overlay intensity view with labelled image view with custom transparency	absolutely required	yes	?	yes	yes
Supports DVID image service	absolutely required	yes	no	no	no
Support Nifti display via http	absolutely required	yes	?	n/a	?
Community uptake	high	Active adoption by several communities in the field: The openconnectome project (neurodata.io) has forked it. The EM / fruitfly community is using it. The developers of ilastik mentioned that it would be their favorite choice to comply with.	None yet. The software is in a young development stage, and not yet very stable under rotations / oblique slicing.	High in the neuroanatomy community. Several linked software projects, e.g. microdraw.	Very high in the neuroimaging community (MRI, DTI)
Efficiency displaying cross-sectional views (obliques)	high	high	potentially high	n/a	n/a
Efficiency loading canonical 2D views	high	high	potentially high	n/a	high
Efficiency when displaying large datasets at a coarse scale	high	High when using the precomputed tile backend. Possibly poor depending on the use of other backends.	potentially high	high	poor
Efficiency when only displaying 2D views	high	moderate, if adjusting chunk sizes to this usecase (allows individual setting of anisotropic chunk sizes per slice view)	?	nearly optimal	poor
Support Gifti mesh format	high	no	no	n/a	yes
Support Nifti display for local files	high	no (could be added)	?	n/a	yes
Support openconnectome ndimage service	high	yes	no	no	no
Supports multi-resolution mesh	moderate	no	no	n/a	no
Support Nifti display for backend-stored files	moderate	yes	?	n/a	yes
supports time series for 2D data	moderate	no	no	no	no
supports time series for 3D data	moderate	no	no	no	no
Suppports URL configuration for data access from any http service	moderate	?	?	yes	no

Figure 18: Feature comparison table for different web-based 3D viewer libraries considered.

5.7.3.2 Component 89-2 Extension of web-based 3D viewer for selecting and displaying a parcellation as a semitransparent overlay

Description of Component: In this extended version, the user will be able to choose a parcellation from a list, which is then displayed as a semi-transparent overlay on top of the original contrast. We assume the parcellation to be given as a labelled (integer) volume dataset.

CDP contribution: This Task directly contributes to CDP3-P2: 3D interactive big data viewer

Progress on Component: Semi-transparent display of parcellations on top of the grey level volume and selection of areas is already possible with the current extended version of Neuroglancer. For the overall user interface with menus, a mock-up has been designed which is at the time of writing under discussion with key users in different SPs for early user feedback (Figure 19).

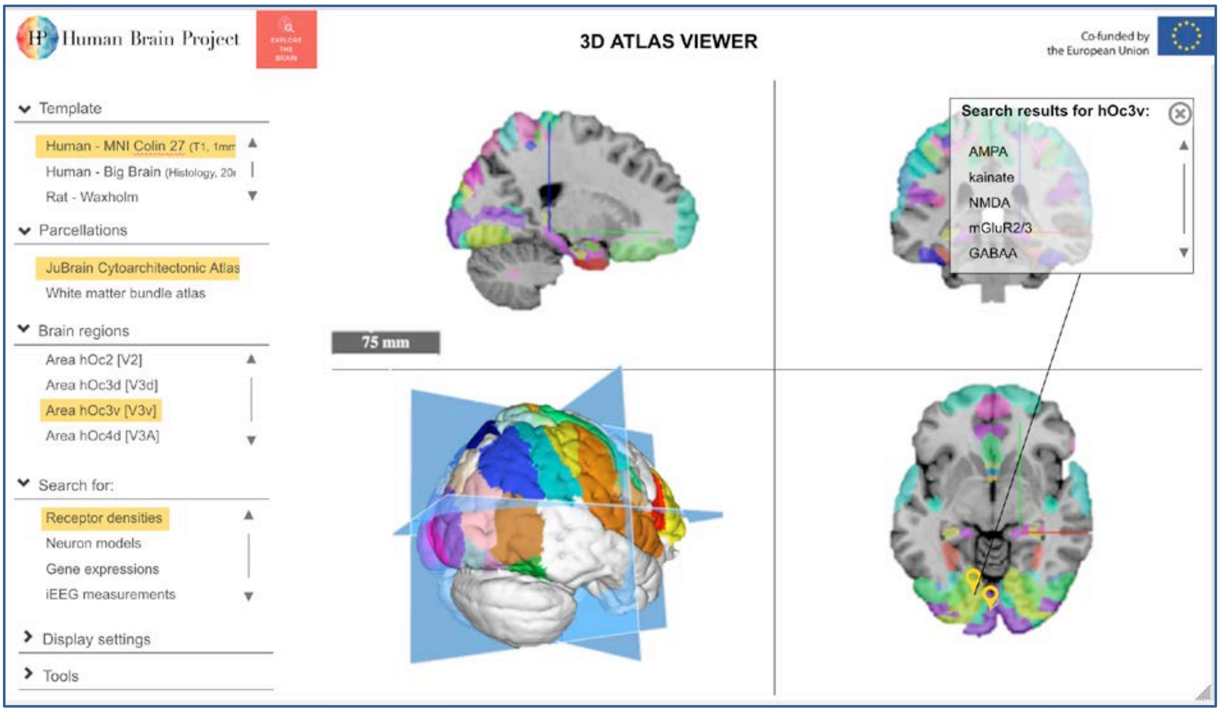

Figure 19: Mock-up for an extended 3D atlas viewer interface based on Neuroglancer, with a menu structure for selecting template spaces, parcellations, and brain regions.

5.7.3.3 Component 90-3 Extension of web-based 3D viewer to perform search queries by interactive selection of a rectangular region of interest

Description: In this extended version, the user should be able to interactively select a rectangular 3D region of interest, and trigger a spatial search for a specific datatype on the search index. The datatype will be preselected, and the search results will be provided as a list of URLs similar to the NIP search mask. This assumes that a spatial search API is available, otherwise a simple mock-up service will be used as a temporary replacement.

CDP contribution: This Task directly contributes to CDP3-P2: 3D interactive big data viewer.

Progress on Component: Not yet started. Connection to NIP services will be the last step of the implementations in SGA1, and requires the availability of HBP image services (T5.5.4), complete curation of the respective reference atlases into the metadata database, and availability of a mature API to the metadata database and search functionality.

5.7.3.4 Component 90-4 Integration of the web-based viewer to the NIP backend database (“KnowledgeGraph”) and HBP federated storage

NOTE: The title and description for Component 90-4 has been updated to more precisely explain how the web-based 3D viewer will be integrated in the NIP.

Description of Component: This Component will connect the 3D viewer to the SP5 and SP7 infrastructure services for 1) retrieving metadata from the KG and 2) retrieving image data from the HBP image service or http-hosted files.

CDP to which Component contributes: This Task directly contributes to CDP3-P2: 3D interactive big data viewer.

Progress on Component: Not yet started. Connection to NIP services will be the last step of the implementations in SGA1, and requires the availability of HBP image services (T5.5.4), complete curation of the respective reference atlases into the metadata database, and availability of a mature API to the metadata database and search functionality.

5.8 T5.4.4 – Integration of Neuroimaging Tools

5.8.1 *Key Personnel*

Task Leader: Simon EICKHOFF (JUELICH)

5.8.2 *SGA1 DoA Goals*

Provide the preprocessing pipelines necessary to bring large volumes or raw MRI-data into a format that is useful for further analysis.

Make neuroimaging tools available on an HCP environment, integrating them with the spatially anchored datasets and the KG.

Using these tools on the datasets aligned in WP5.3 will allow production of new derivative data adding information on structure, function and connectivity to the parcellations available in the aligned brain atlases, providing information on inter-individual variability and relationship to behaviour.

Provide query tools that allow researchers in the neuroscience SPs to address their specific questions while at the same time registering the output into the KG.

This Task will create the necessary software environment and evaluate different pipeline solutions for data preprocessing and initial analytical questions. These first Use Cases will help to establish the general framework serving a dual purpose. On the one hand, neuroscience researchers can use the HBP resources to address individual scientific questions; on the other, the derivatives will also be integrated into the NIP.

5.8.3 *Component Progress*

5.8.3.1 Component 91-1 Preprocessing pipeline for raw neuroimaging data

Description of Component: Given that most data shared by institutions around the world are provided as raw [Digital Imaging and Communications in Medicine (DICOM) or NIfTI] data and that processing should be standardised across datasets, we will set up the necessary pipelines to process these into datasets that are ready to be used by researchers in the HBP and beyond. This includes implementing pipelines that cover surface- and volume-based structural imaging, resting- and Task-state fMRI as well as diffusion-weighted imaging (both standard and HARDI).

Progress on Component: Work on preprocessing pipelines for multi-modal MRI data is progressing largely as planned with faster than anticipated progress on several tools that can be compiled as standalone functions in a HPC environment and unforeseen progress in the addition of GPU acceleration for several functions. In turn, several other Components that rely on compiled MATLAB-code, third-party libraries or additional environments, progress has been slightly slower than planned. We expect the implementation of the complete processing pipeline and its initial application to at least two large-scale datasets of ~1,000 subjects to be finished at the end of SGA1.

5.8.3.2 Component 92-2 Implementation of atlas-based analysis pipelines

Description of Component: We will implement the HPC based extraction of features from the Human Connectome data for regions defined by one of the HBP atlases. We expect the first instalment of this to be not yet fully linked to other datasets through the KG, giving on-going developments in other Tasks, but will set up the necessary infrastructure that can be linked with the rest of the NIP in SGA2.

Progress on Component: After some difficulties in hiring suitable personnel for this work, we have now a first prototype software for extracting and analysing structural features in a framework that will be compatible with the HBP-HPC resources. We expect this work to be completed in time and may be able to additionally add data on intracortical myelin. In addition, we should be able to set up pipelines for the extraction of additional features based on resting-state functional connectivity through a new in-kind contribution of 8 person months starting this summer.

6. WP5.5 Community-Driven Neuroinformatics Platform and Infrastructure Operations

See Figure 20.

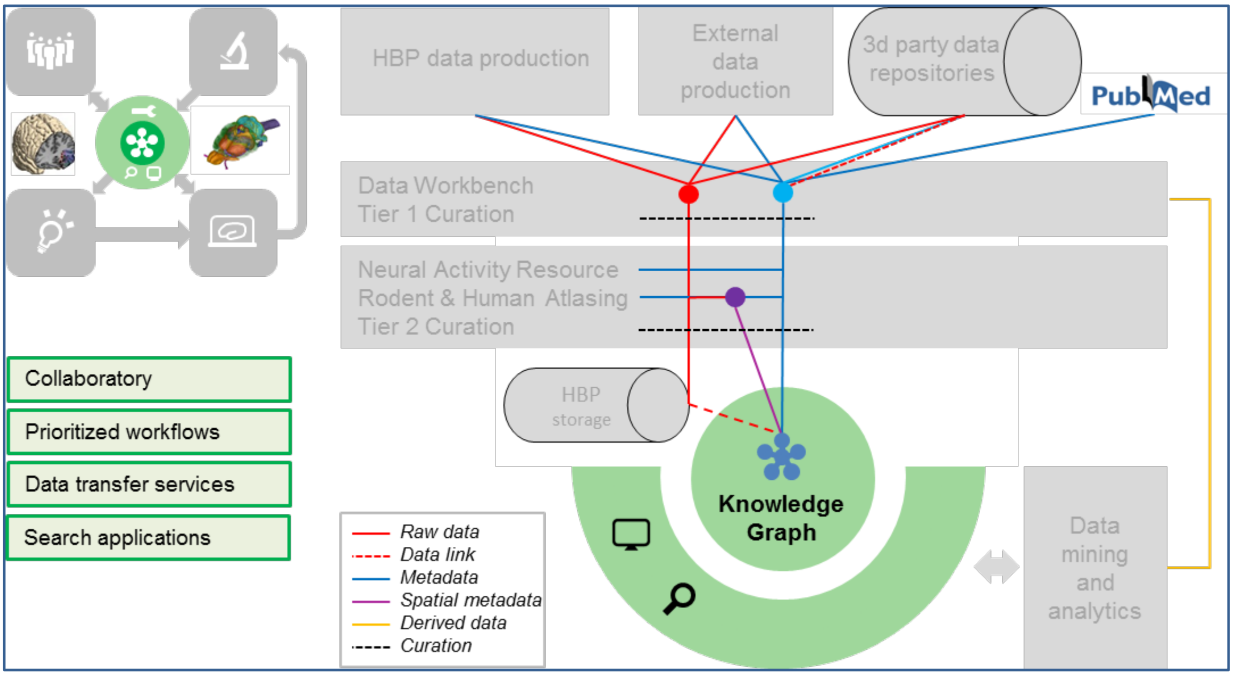

Figure 20: Schematic overview of contributions from WP5.5, highlighted in green.

6.1 Key Personnel

WP Leader: Jeff MULLER (EPFL)

Task Leaders: Jeff MULLER (EPFL)

Anastasia AILAMAKI (EPFL)

6.2 Work Package Leader's Overview

To follow up on the changes in the new work plan, and to replace the team previously responsible for major parts of this WP, hiring is ongoing and progressing along expected lines.

A key advance has occurred since the hiring of three new staff members in February and March. This team is involved in work in T5.5.3, T5.5.5 and T5.5.6 and is heavily involved in transition to FENIX based hosting of the Collaboratory. This collaboration has been progressing well with strong support from the CSCS infrastructure support teams. It is expected that 80% of the Collaboratory production services will be running on CSCS Openstack base infrastructure by the end of April 2017.

Another notable development has been planning and evaluation activities around viewers have been progressing rapidly in strong collaboration with WP5.4, WP5.3 and SP7.

Over the first 12 months of the SGA1 only T5.5.5 and T5.5.7 have been operating. As a result, progress in WP5.5 is concentrated in T5.5.5 and T5.5.7 with planning, evaluation and documentation activities happening in other Tasks. This is in line with the finalised D5.6.1 delivered at the end of January 2017, but is important to consider when reviewing the impact.

Finally, despite delays in hiring due to the approval of Deliverables, the collaboration between WP5.5 and other WPs in SP5 is working well and common objectives are clear.

Formulation of project-wide strategy is being led by the strong involvement of WP5.5 in public website strategic direction and implementation as well as infrastructure strategy development.

6.3 Priorities for the remainder of the phase

The main priorities in the remainder of the phase are:

- Implement KG and KG search based on the outcomes of the KG evaluation report, which has been completed. This will be in strong collaboration with T5.4.1 to ensure that curation of data can proceed regardless of service availability.
- Deploy production image services focusing first on delivery of rich reference atlases and subsequently on “working” dataset services needed for production of refinements to the reference atlases and derived atlas reference datasets coming from WP5.2, WP5.3, and WP5.6.
- Improve capacity and performance of the Collaboratory. As load and user activity have been growing, performance bottlenecks are being identified and addressed. Notably, several MOOCs are planned for 2017, which will require further capacity tests, capacity planning, and capacity enhancement.

6.4 Milestones

See Table 5.

Table 5: Milestones for WP5.5 — Community-Driven Neuroinformatics Platform and Infrastructure Operations

MS No.	Milestone Name	Lead Partner	Task(s) involved	Month Due	Month of Completion	Comments
MS5.5.1	List of KPIs for internal and external user adoption of Products/services offered by WP5.5	P1 (EPFL)	T5.5.5	M10	*M10*	Revised and finalised by SP Leader. Approved by Task/WP leaders. Minor adjustments of KPIs for internal SP use will be considered based on accumulated experience during the project period. A pre-verification version of this Milestone was delivered to the European Commission together with the D5.6.1 Resubmission in January 2017. The full list of KPIs is available for the HBP SGA1 consortium through the Milestone Report Repository and can also be found in Appendix 1. In this report, we include the M18 report on KPIs, with comments on adjustments of targets for SGA1
MS5.5.3	*KnowledgeGraph Evaluation Report*	P1 (EPFL)	T5.5.1	M12	*M20*	Achievement date: 08 November 2017. The Components providing input to this Milestone are the RUP KnowledgeGraph and KnowledgeGraph Search components. Verification will be made by availability and review of the planned KnowledgeGraph Evaluation report, in line with the response to the review of the DPIT work plan, submitted to the EC on 31 December 2016. The achievement of this Milestone is slightly delayed due to extended discussions around the open source licensing of input RUP components described above.

						This may slightly delay progress on the KnowledgeGraph, but the expected economies of scale in a strong, open source, multi-partner KnowledgeGraph component is considered worth the small risk. The full report is available for the HBP SGA1 consortium through the Milestone Report Repository. *https://emdesk.humanbrainproject.eu/shared/5a12af89ab6ca-66266950eafa52f81cc3dcba808fe7f2* and https://emdesk.humanbrainproject.eu/shared/5a12af9733f83-ddadd7fc23c389ab29f456c3eb5e82a0
MS5.5.5	Program Evaluation Review Technique (PERT) chart showing global HBP data-related work breakdown structure and prioritised workflows for SP5 in interaction with the entire HBP	P1 (EPFL)	T5.8.1 T11.2.4	M12	*M14*	The principal parts of the PERT charts are defined and verified by WP and Task leaders. The PERT chart is available for the HBP SGA1 consortium through the Milestone Report Repository. https://emdesk.humanbrainproject.eu/shared/59a661b2ac2a3-68b8b0fe9d7e1f0a525dc85f638dd936. At the time of submission of D5.8.2, milestones were uploaded in an HBP repository. Unfortunately, reviewers did not have access to this repository due to technical limitations

6.5 T5.5.1 – Development and Maintenance of KnowledgeGraph

6.5.1 *Key Personnel*

Task Leader: Jeff MULLER (EPFL)

6.5.2 *SGA1 DoA Goals*

Transition from the work done by the Blue Brain Project in the RUP phase to a more flexible and open approach to metadata stores supporting more widespread curation efforts for data inside HBP and beyond.

6.5.3 *Component Progress*

6.5.3.1 Component 93-1 KnowledgeGraph Service

Description of Component: The KG Service is a REST service with a database (relational or NoSQL) storing-provenance metadata as well as spatial anchoring of datasets into atlas templates. This database will be used for data ingest as well as non-interactive queries (the latter may be restricted to NIP internal users and services).

CDP to which Component contributes: All CDPs.

Progress on Component: Over recent months, key technical evaluations related to the KG have been progressing smoothly. The KG evaluation report has been completed..

6.5.3.2 Component 93-2 KnowledgeGraph Elastic Search Index Service

Description of Component: The KG Elastic Search Index Service as a clustered, scalable search index for queries where interactive performance is required. This Task is responsible for installation, optimisation and maintenance of the Elastic Search cluster required to support interactive queries. This service is private and is only publicly accessible through Product 95-1- KG Search API.

Progress on Component: No progress due to a need to complete the evaluation report prior to working on deployment or development of new functionality. Work will commence after submission of the KG Evaluation Report.

6.5.3.3 Component 93-3 KnowledgeGraph Evaluation Report

Description of Component: This report will address:

1) architectural flexibility of the RUP KG;
2) progress towards transferring operations ownership to the HBP;
3) a memorandum of understanding regarding shared development between HBP and BBP.

Progress on Component: This Component is the subject of MS5.5.3 and is reported there.

In lieu of the MS5.5.3 report: Work on this report is nearing completion but has been delayed due to open source licensing negotiations which needed to be concluded to complete the evaluation report.

6.5.3.4 Component 93-4 KnowledgeGraph Federation Architecture Report

Description of Component: Federating a system like the KG is no trivial matter, especially at a global scale. This report will be the outcome of a prototyping and evaluation activity undertaken by T5.5.1 and T5.5.2 along with database expertise from T5.5.8. This report will set the architecture and roadmap for development of the federated KG based on concrete Use Cases and detailed technical assessments. This is crucial groundwork for implementation of such a system in future SGAs when a federated KG is deemed strategically important.

Progress on Component: Work on this Component will start in M18 for delivery in M24. The Component is the subject of MS5.5.4 and will be reported under this heading when due.

6.5.3.5 Component 94-2 KnowledgeGraph Python API

Description of Component: Many of the add-on services such as the registration UI will be built with Python. In addition, Python is well supported by existing Collaboratory Jupyter notebooks and other neuroscientific workflows. This work will aim to make integration of provenance registration into scientific workflows and web applications simple and straightforward to increase adoption of the HBP KG inside and outside the HBP. This library will be open-sourced to facilitate community building.

CDP to which Component contributes: All CDPs.

Progress on Component: As planned, work on this Component has been restricted to evaluation of the RUP KG produced in the Blue Brain Project, as described in the KG Evaluation Criteria, described in the responses to the DPIT review.

6.6 T5.5.2 - Search Application

6.6.1 *Key Personnel*

Task Leader: Jeff MULLER (EPFL)

6.6.2 *SGA1 DoA Goals*

Provide REST APIs and Python client libraries for a metadata search index service with a focus on metadata coming from KG, Collaboratory, Software Catalog and selected third party repositories.

Intuitive search UI for the data stored by repositories throughout the HBP.

6.6.3 *Component Progress*

6.6.3.1 Component 95-1 KG Search API

Completed in RUP.

6.6.3.2 *Component* 96-2 Collaboratory integrated Search UI

Description of Component: A Collaboratory Integrated UI to allow search for entities in the KG Search API and to navigate to locations where the user can find more information on each of the entities returned by the search. Significant effort will be expended to ensure that the UI is usable and productive for power users and non-power users alike.

CDP to which Component contributes: All CDPs.

Progress on Component: As planned, work on this Component has been restricted to evaluation of the RUP KG produced in the Blue Brain Project, as described in the KG Evaluation Criteria, described in the responses to the DPIT review.

6.6.3.3 Component 97-3 KG Search Python API

Description of Component: To facilitate automation and scripted workflows using the KG, this Component will provide a Production grade Python client API for the KG Search API. This API is expected to see heavy use in the Collaboratory's Jupyter notebooks to allow interactive programmatic exploration of the data visible through the KG Search API.

CDP to which Component contributes: All CDPs.

Progress on Component: As planned, work on this Component has been restricted to evaluation of the RUP KG produced in the Blue Brain Project, as described in the KG Evaluation Criteria, described in the responses to the DPIT review.

6.6.3.4 Component 98-4 KG Search Indexer

Description of Component: To populate the KG, a collection of indexers will be delivered for the following data sources including, but not limited to:

- Collab service;
- Collab Storage service;
- KG Service.

CDP to which Component contributes: All CDPs, indirectly.

Progress on Component: As planned, work on this Component has been restricted to evaluation of the RUP KG produced in the Blue Brain Project, as described in the KG Evaluation Criteria, described in the responses to the DPIT review.

6.7 T5.5.3 - Development and Maintenance of Web Front End to the Platform Services

6.7.1 *Key Personnel*

Task Leader: Jeff MULLER (EPFL)

6.7.2 *SGA1 DoA Goals*

Provide frontend and back-end web development to ensure high-quality customisation and integration of various solutions developed and deployed throughout SP5 into the Collaboratory.

In addition, this Task will work to ensure that SP5 has strong linkages between clear content on the redesigned public facing website and the activities in the HBP Collaboratory. This effort is essential to highlight the outputs of SP5 and the SPs which rely on SP5 to make their data available to a wider audience; SP1, SP2 and SP3.

6.7.3 *Component Progress*

6.7.3.1 Component 99-1 Web site

Description of Component: A Neuroinformatics public-facing website with anonymous access areas will be rebuilt to ensure that documentation, Software Catalogs, and applications to search for data, as well as atlas viewers, can be accessed without registering for an HBP Collaboratory account. Special care will be needed to integrate with the public-facing https://www.humanbrainproject.eu and ensure easy access to the NIP for visitors who are new to the HBP.

Progress on Component: On 20 April 2017, the new HBP website was released, including a tab entitled Explore the brain. This tab is the new public access website for the NIP.

Links: http://www.humanbrainproject.eu/en/explore-the-brain

6.7.3.2 Component 100-2 UI development support

Description of Component: For all web applications developed in SP5. This Task will provide UI development support to ensure a consistent look and feel and to ensure usability across SP5's various web interfaces.

Progress on Component: This Task has been supporting the redesign of the public-facing website in line with a revamped communication and engagement strategy. The first version of the public website was released on 20 April and highlights notable outputs in SP5 like the Neuroglancer based atlas viewers along with those developments in other SPs. An extensive process of stakeholder engagement has been ongoing throughout the redesign project to ensure that the SP5 site is aligned with the roadmap for the respective SP5 activities.

6.8 T5.5.4 - Data and Image Services

6.8.1 *Key Personnel*

Task Leader: Jeff MULLER (EPFL)

6.8.2 *SGA1 DoA Goals*

Perform the work required to provide a standard image service for all atlasing viewer Use Cases.

The image service will deliver volumetric, image and time-series data (2-4D data sets) and geometric data needed to allow T5.4.2 and T5.4.3 to provide Productive atlas exploration tools. The image service will be accessible via REST API and will be deployed at data centres hosting large atlasing datasets.

The Image service will also deliver image, volumetric or geometry datasets registered in the reference atlases datasets to the image viewers. An effort will be made to integrate this service into atlas building tools and workflows, but this is considered a prototype effort in SGA1 and will be subject to data volumes processed by the atlasing workflows, as well as data locality optimisations available in the implemented image service.

6.8.3 *Component Progress*

6.8.3.1 Component 101-1 Image Service

Description of Component: REST APIs delivering rich URL-indexed image, voxel, time series and geometric data. Transitioning this service to a high TRL service in SGA1 will require delivering a well-documented specification for the API provided by the service. This will also be crucial to allowing groups outside the HBP to build their own clients and to allow more loosely coupled development between the Image Service and the HBP atlases.

Progress on Component: The requirements for this API are covered in the documentation of Use Cases and requirements analysis performed preparation of D5.8.1. Evaluation of various services has been undertaken with the intention of early deployment of reference atlases in such a way that they can be made interactively explorable directly from the HBP public website.

6.8.3.2 Component 102-2 Python Image Service Client

Description of Component: Python client APIs for the Image service to enable workflow scripting and analytic Use Cases. This Component depends in part on a stable, documented rest API as described in Image Service Component of this Task.

CDP to which Component contributes: CDP1 and CDP3.

Progress on Component: The requirements for this API are covered in the documentation of Use Cases and requirements analysis performed preparation of D5.8.1. It is important to consider that this Component will only be developed to prototype maturity levels in SGA1 and development work is not expected to start until late in SGA1.

6.8.3.3 Component 103-3 Data Analytics Service

Description of Component: REST APIs delivering rich URL-indexed image, voxel, time series and geometric data. This was provided by the Voxel Brain API in the Ramp-Up Phase but will be developed in part with an external contractor in the SGA1.

CDP to which Component contributes: CDP1 and CDP3.

Progress on Component: The requirements for this service are covered in the documentation of Use Cases and requirements analysis performed preparation of D5.8.1. It is important to consider that this Component will only be developed to prototype maturity levels in SGA1 and development work is not expected to start until late in SGA1.

6.8.3.4 Component 104-4 Interactive Data Analytics Notebook Service

Description of Component: A prototype Big Data image query and processing system on top of Apache Spark. The target will be to prototype work needed for atlas building and rich analytic query processing in SGA2.

CDP to which Component contributes: CDP1 and CDP3.

Progress on Component: The requirements for this service are covered in the documentation of Use Cases and requirements analysis performed preparation of D5.8.1. It is important to consider that this Component will only be developed to prototype maturity levels in SGA1 and development work is not expected to start until late in SGA1.

6.9 T5.5.5 – Collaboratory

6.9.1 *Key Personnel*

Task Leader: Jeff MULLER (EPFL)

6.9.2 *SGA1 DoA Goals*

Bring largely independent collections of software and services together into a productive ecosystem for scientific projects; provide basic set of common web services and a central portal for the Collaboratory and Platforms.

One of the key functions requested by Platform developers throughout the HBP is a central web-accessible storage location with REST API access and a web UI. This storage is intended for collaboration and scientific workflow support. This Task is responsible for feature development, maintenance and extension of this capability in response to feature requests and service usage growth. Storage capacity will be provided via FENIX/SP7.

Another key function will be the Collaboratory integration of Jupyter notebooks. Jupyter notebooks are a collaboration technology that provides a web-based literate programming tool with execution of Python workloads (other languages are supported, but will not be the focus of this Task). These are used today in the HBP for documentation, interactive instruction, automation and exploration. This Task will seamlessly integrate Jupyter notebooks as Collaboratory Apps and viewers. This Task also configures computing resources for running the execution kernel of the Jupyter notebooks.

6.9.3 *Component Progress*

NOTE 1: This Task report includes two new Components (105-1b and 105-1c) which were not explicitly listed in the work plan, "Collaboratory Storage IU" and "Collaboratory Jupyter Notebook". Both Components were existing services in the Collaboratory ecosystem and are reported on here to highlight areas of major work in the Collaboratory task and allow reporting in a more consistent and complete manner.

NOTE 2: The description of Component 105-1 contains a correction of a mistake in the D5.6.1 work plan which replicated a section of the Image Service component description.

6.9.3.1 Component 105-1 Collaboratory

Description of Component: New description. The following apps and services are the basis for the Collaboratory ecosystem and supported for HBP internal and external developers.

- HBP Identity - OpenID Connect based AAI infrastructure used by services throughout the Project
- Collaboratory
- Chat
  - Navigation
  - Wiki + live documents
  - Documentation hosting service
- Team Profiles
- Software and service catalogue for software and service Components developed and deployed throughout the HBP infrastructure

- Additional capabilities may be added in form of independent apps

Progress on Component: The Collaboratory has been updated throughout the first 12 months to improve the user experience and fix outstanding issues in various services. Notable improvements include:

- upgrades to so-called “Live documents” a popular collaborative document editor;
- numerous bug fixes;
- improvements in the presentation of public and private collabs;
- improvements in the presence section of the Collaboration panel;
- improvements in responsive design behaviour of the Collaboratory;
- upgrades, maintenance and deployment of the HBP Identity service;
- addition of a Collab overview which tracks activities in a given collab.

Links: https://collab.humanbrainproject.eu/

6.9.3.2 Component 105-1b Collaboratory Storage UI

Description of Component: Web UI for the Collab Storage service and the UNICORE Storage REST service. Also provides AngularJS components for use in other AngularJS applications.

Progress on Component: This Component was developed with an iRODS dependency in the RUP which proved problematic for performance optimisation and maintenance when it was put into operation. Development thus far in SGA1 has been focused on removing the iRODS dependency in favour of a direct connection to a relational database management system for metadata and a direct connection to S3. This is done in a managed fashion with a well-defined migration path from the old “v0” REST API to an optimised “v1” API. The foundation work on the resulting v1 has been completed and deployed. Significant performance improvements in the Storage UI will be visible when it has been migrated to the new v1 API. This is expected to happen by the end of April 2017.

Links: Visit the Storage in any user Collab. Create a collab here and visit the Storage navitem to see how it works. https://collab.humanbrainproject.eu/#/collab/19/nav/179

6.9.3.3 Component 105-1c Collaboratory Jupyter Notebook

Description of Component: A web based Jupyter Python notebook for interactive usage. Integrated with the Collaboratory and the Collaboratory Storage service to provide a key service for a more seamless collaborative workbench for Python based scientific collaboration.

CDP to which Component contributes: CDP2 and CDP4.

Progress on Component: Numerous updates have been applied to the Jupyter notebook service over the first 12 months. Upgrades from the main open source branch have been integrated as they have become available. Additionally, the management of the Jupyter kernel service lifecycle has been improved to improve resource utilisation and reduce management burden. Finally, UI improvements have been added to ensure that users are notified if they leave a page with a modified notebook. These changes are necessary because the Jupyter notebook is primarily designed for single-user desktop deployments, not for a large-scale multi-user deployment in an environment such as the Collaboratory. Finally, a base library of neuroscience software is regularly updated to ensure that users have access to software necessary for their workflows.

Links: For information on how to use the Jupyter notebook in the HBP Collaboratory and for Release notes on past releases visit:

https://collab.humanbrainproject.eu/#/collab/509/nav/11881

6.10 T5.5.6 – Platform Administration, Operation and Validation

6.10.1 *Key Personnel*

Task Leader: Jeff MULLER (EPFL)

6.10.2 *SGA1 DoA Goals*

Operate and support the NIP services in Production and provide services and support to improve SP5's ability to deliver those reliable services into the Production environment in a predictable way.

To accomplish this goal, this Task will provide DevOps services to the Platform development teams in SP5. These services include, but are not limited to:

1) support for SP5 use of VM base image configuration and Docker container infrastructure;
2) service monitoring infrastructure for SP5 offered services;
3) security reviews and security patches on base software;
4) common database services with security and backup policies established;
5) support for continuous integration services related to software release and deployment;
6) architectural recommendations to minimise cost and maximise scalability and service resilience;
7) Platform validation with automatic collection of KPIs where possible.

6.10.3 *Component Progress*

6.10.3.1 Component 106-1 HBP Standard Deployment service

Description of Component: PaaS offering to provision, configure and deploy VMs at the respective sites with optimised access to appropriate Data Services (shared between T5.5.6 and T5.7.2).

CDP to which Component contributes: All CDPs, foundation activity for the majority of SP5 services.

Progress on Component: This is one of the tasks which was part of the D5.6.1 work plan and as such staff were only hired on 1 February 2017. In spite of this, progress has been extensive. Initial versions Platform-as-a-Service offerings for logging and automated collection of metrics have been set up at the FENIX site in CSCS. Base Ansible configurations have been established to ease migration of services to CSCS and test deployments of selected services have been completed. Production deployment activities started 30 March 2017 and it is expected that 80% of production Collaboratory services will be migrated to CSCS by April 2017.

6.11 T5.5.7 – Spatial Search Application

6.11.1 *Key Personnel*

Task Leader: Anastasia AILAMAKI (EPFL)

6.11.2 *SGA1 DoA Goals*

The goal of this Task is to provide (ad hoc) spatial search within KG: enable users to retrieve every dataset (or its metadata) that have its spatial data (location/position) intersecting a region of interest. The region of interest can be defined as a point or bounding box (in 3D space). The Task includes developing a spatial index component within the KG and corresponding search (REST) API.

6.11.3 *Component Progress*

6.11.3.1 Component 111-1 Spatial Search API

Description of Component: REST API to perform ad-hoc spatial queries in KG.

CDP to which Component contributes: CDP1 and CDP3.

Progress on Component: No progress on the REST API Component. The focus has been on integration of enhanced spatial query support in the Component in 6.11.3.2. In later phases of development, this Task is expected to work closely with the T5.5.2 to ensure that API development in T5.5.2 is aligned with spatial capabilities developed in this Task.

6.11.3.2 Component 111-2 Spatial Index for KnowledgeGraph

Description of Component: Spatial index (plugin or a separate component in KG) responsible for spatial data organisation.

CDP to which Component contributes: CDP1 and CDP3.

Progress on Component: Ongoing. Efforts are focused on developing a spatial indexing support for a 3D Cartesian/Euclidean space. A design decision has been made to develop spatial querying tightly integrated with ElasticSearch, which is used in KG and specifically in T5.5.2 for text/keyword search. Since it is expected that NIP users will perform searches based on both text and spatial information, the resulting search components will be tightly integrated and thus can offer a better performance. The internals of ElasticSearch, particularly its core component Lucene, have been investigated and (slightly) delayed (but on track) developments have started. Also, we have observed that organising neuroscience datasets into traditional minimum bounding boxes (MBBs) results in a very poor approximation of the underlying spatial data. Therefore, we have investigated a number of alternative approaches, including our novel proposal of clipped bounding box (CBB).

7. WP5.6 Data Mining and Analysis Neuroinformatics Capabilities

See Figure 21.

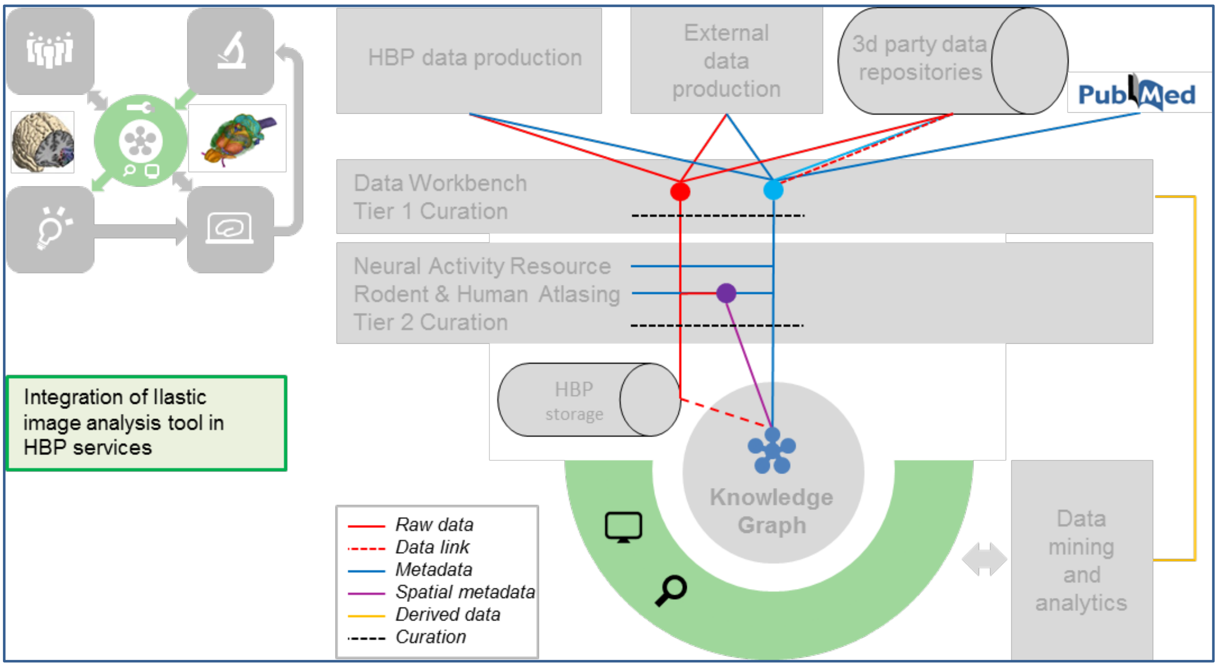

Figure 21: Schematic overview of contributions from WP5.6, highlighted in green

7.1 Key Personnel

WP Leader: Fred HAMPRECHT (UHEI)

Task Leaders: Anna KRESHUK (UHEI)

Pascal FUA (EPFL)

7.2 Work Package Leader's Overview

Following acceptance of the revised work plan in January, both developer positions are filled, and multiple back-end architecture options have already been explored. The first version of the segmentation algorithm for T5.6.3 has already been developed despite the lack of manually annotated data.

In the original plan, our efforts for M1-M12 were concentrated on necessary ilastik architecture changes, while integration with HBP tools was planned for M13-M24, after the viewer and image service APIs are stabilised. However, because of delays in work plan acceptance and recruitment, both will have to be done at the same time.

This work will serve as the foundation for all activity in this WP in SGA1 and SGA2.

7.3 Priorities for the remainder of the phase

The main priorities in the remainder phase will be to:

- integrate NIP image service as a data source into ilastik to enable HBP users to experiment with ilastik algorithms on data from HBP storage;
- break the monolithic ilastik application into viewer, controller and computational back-end;
- make it possible to run the back-end as a service;

- allow control of the back-end from the NIP web-based 2D/3D viewer, including running ilastik on HBP computational resources;
- develop Javascript control widgets to be integrated with the viewer;
- develop session handling tools to display logs, transmit error messages, stop computations or continue them in batch mode;
- allow ilastik results and interactively provided user annotations to be stored;
- integrate new algorithms for segmentation and counting of cellular structures.

7.4 Milestones

See Table 6.

Table 6: Milestones for WP5.6 — Data Mining and Analysis Neuroinformatics Capabilities

MS No.	Milestone Name	Lead Partner	Task(s) involved	Month Due	Month of Completion	Comments
MS5.6.1	List of KPIs for internal and external user adoption of Products/services offered by WP5.6	P47 (UHEI)	T5.6.1-3	M10	M10	Revised and finalised by SP Leader. Approved by Task/WP leaders. Minor adjustments of KPIs for internal SP use will be considered based on accumulated experience during the project period. A pre-verification version of this Milestone was delivered to the European Commission together with the D5.6.1 Resubmission in January 2017. The full list of KPIs is available for the HBP SGA1 consortium through the Milestone Report Repository and can also be found in Appendix 1. In this report, we include the M18 report on KPIs, with comments on adjustments of targets for SGA1

7.5 T5.6.1 - Interactive Feature Classification and Extraction for Spatial Images

7.5.1 *Key Personnel*

Task Leader: Anna KRESHUK (UHEI)

7.5.2 *SGA1 DoA Goals*

Enable users to run ilastik via the NIP

For batch prediction, the NIP will offer the user a choice of already existing classifiers for a given dataset and allow for uploading a new classifier. The algorithm will then process the dataset remotely, near the location of the data, using and building on the resources and tools for parallel processing, provided by SP7 and FENIX. For the interactive prediction, we will add the option to replace the “volumina” viewer of ilastik by the web- based viewer, developed as a part of NIP (Tasks 5.4.2 and 5.4.3).

7.5.3 *Component Progress*

7.5.3.1 Component 109-1 Connection of ilastik to HBP 2D and 3D viewers

Description of Component: Enable the use of ilastik through the HBP image viewers.

Progress on Component:

- Recruited a developer for the task, starting from 15.02.2017
- Defined preliminary architecture (Figure 22) and API for viewer/back-end communication
- Built an experimental prototype for the client/server setup
- Main difficulty: interactive update of the viewer layers and invalidation of the viewer caches on parameter changes

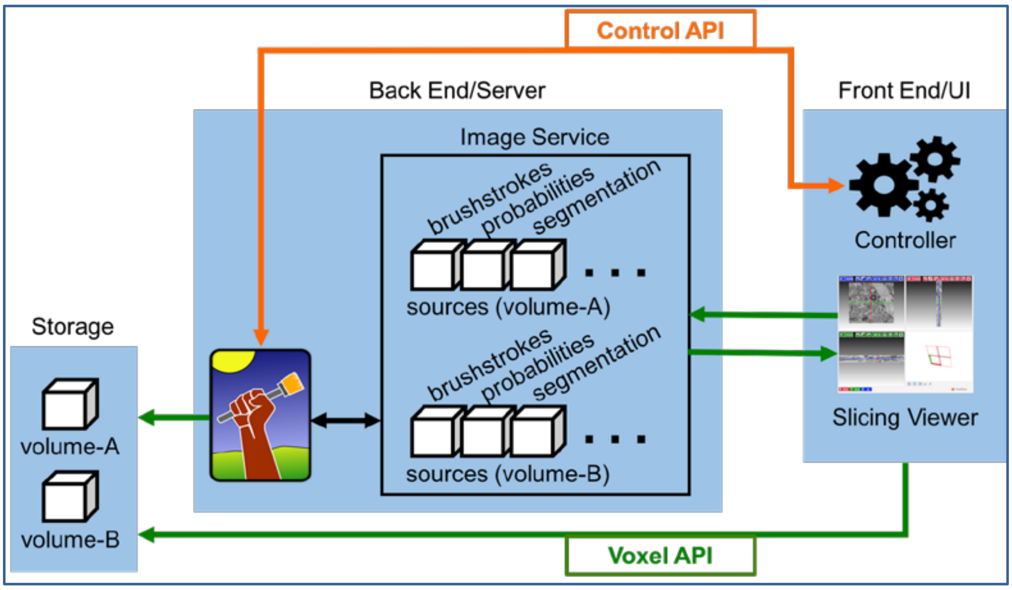

Figure 22: Principle architecture for separation of ilastik into a server-based back-end and a client-side, web-based front-end.

Orange arrows indicate communication via the control API, green arrows indicate communication via the voxel API

7.6 T5.6.2 - Workflow for Populating Brain Atlases with Features, Automatically Extracted by ilastik

7.6.1 *Key Personnel*

Task Leader: Anna KRESHUK (UHEI)

7.6.2 *SGA1 DoA Goals*

Connect ilastik with the image service of T5.5.4 and allow it to use resources provided by SP7 and FENIX

This includes correct handling of metadata and spatial anchors and their transfer to the derived datasets (ilastik results) and user annotations, which implies integration with the KG of T5.5.1. For efficient use of SP7 resource, ilastik should be able to run as a service, controlled through either through the viewer (see T5.6.1) or through the Collaboratory.

7.6.3 *Component Progress*

7.6.3.1 Component 110-1 Connection of ilastik to other HBP services

Description of Component: As a result, ilastik will be able to work with HBP data and to save its own results and the manual user annotations in a way to be used by the related Tasks.

Progress on Component: Recruited a developer for the Task, starting from 1 January 2017. Organised an internal hackathon, where multiple back-end engine prototypes were evaluated with the goal of preserving as much as possible of existing algorithm code and replacing the engine to run as a service.

Following options considered: DASK and PyTorch, ITK, Python with the futures library, Intel Thread Building Blocks. Most promising prototypes further developed now.

Main elements of a workflow for WP5.6 have been developed and agreed on with SP7. A technical Use Case description, serving as a basis for implementations to be made, has been developed in collaboration with SP7 (see Appendix 2).

7.7 T5.6.3 - Integrating Feature Extractors and Classifiers for Neuroscience

7.7.1 *Key Personnel*

Task Leader: Pascal FUA (EPFL)

Other Researcher: Eduard TRULLS (EPFL)

7.7.2 *SGA1 DoA Goals*

Automated detection of cellular structures on electron microscopy (EM) scans is of paramount importance, as the scale of the data produced in HBP makes manual processing simply not feasible. While off-the-shelf solutions are available, the state of the art in visual recognition lies in machine-learning techniques. The goal of this Task is to develop tools based on convolutional neural networks for detection and counting of structures such as cells or synapses, which will be provided as ilastik plugins for HBP workflows and services available through the NIP.

7.7.3 *Component Progress*

7.7.3.1 Component 112-1 SGA1 T5.6.3: Machine-Learning Tools for Segmentation

NOTE: The title of this Product/Component in the D5.6.1 Resubmission was Segmentation tools. Both the title and description of this Component has been modified to adapt to the use of ilastik as a framework under this Task.

Description of Component: We are developing algorithms based on machine-learning techniques to automate segmentation and counting of structures such as cells of synapses. They will be integrated into HBP as ilastik plugins.

Progress on Component: Ongoing. We have developed an initial version of our segmentation algorithm and conducted a small-scale cell-counting experiment on the GAD67-GFP deNeo mouse brains, with promising results. See Figure 23. The main roadblock is the shortage of ground truth annotations, which are critical to train our models on such large datasets (~4 Tbytes). To this end we are collaborating with Graham KNOTT at the BioEM imaging lab at EPFL, who is helping us annotate a larger stack. We are also in the process of identifying more relevant users and datasets for our tools, which proved challenging in the first half of SGA1 given the delayed start for many partners, but now shows steady progress.

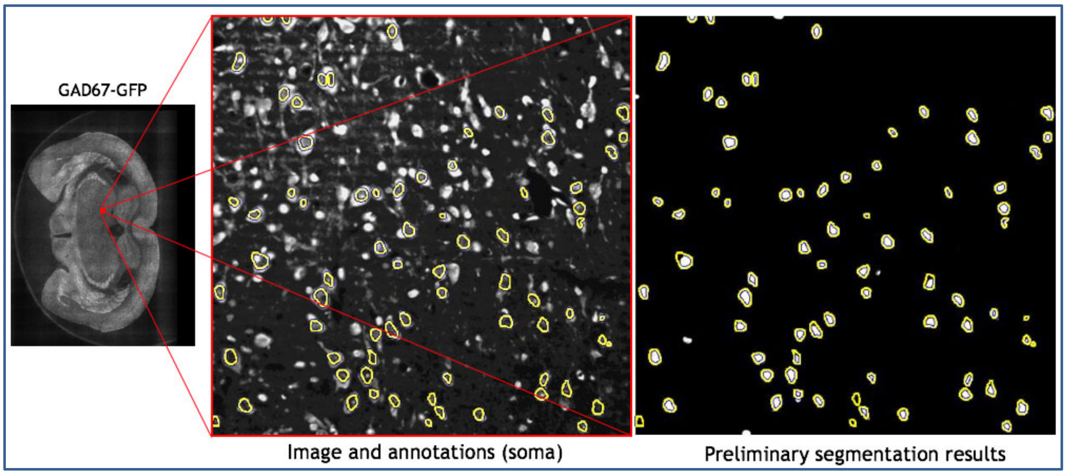

Figure 23: Early segmentation results on the GAD67-GFP deNeo mouse brain.

Integration is also progressing. Our algorithms will be developed as services running on GPU-powered servers which can be accessed remotely using ilastik as a front-end, leveraging its powerful annotation and visualisation capabilities while preserving flexibility on the back-end. We have a functional prototype and have integrated, as a proof of concept, our RUP segmentation algorithm (originally integrated into ESPINA and delivered in RUP Product 112-1). Our services could potentially be accessed in other ways, such as a simple web interface, for batch processing with pre-trained models, or IPython notebooks, for power users. We will explore these possibilities according to the needs of early adopters.

8. WP5.7 Tools and Curation for Integrated Parallelised Analysis of Activity Data

See Figure 24.

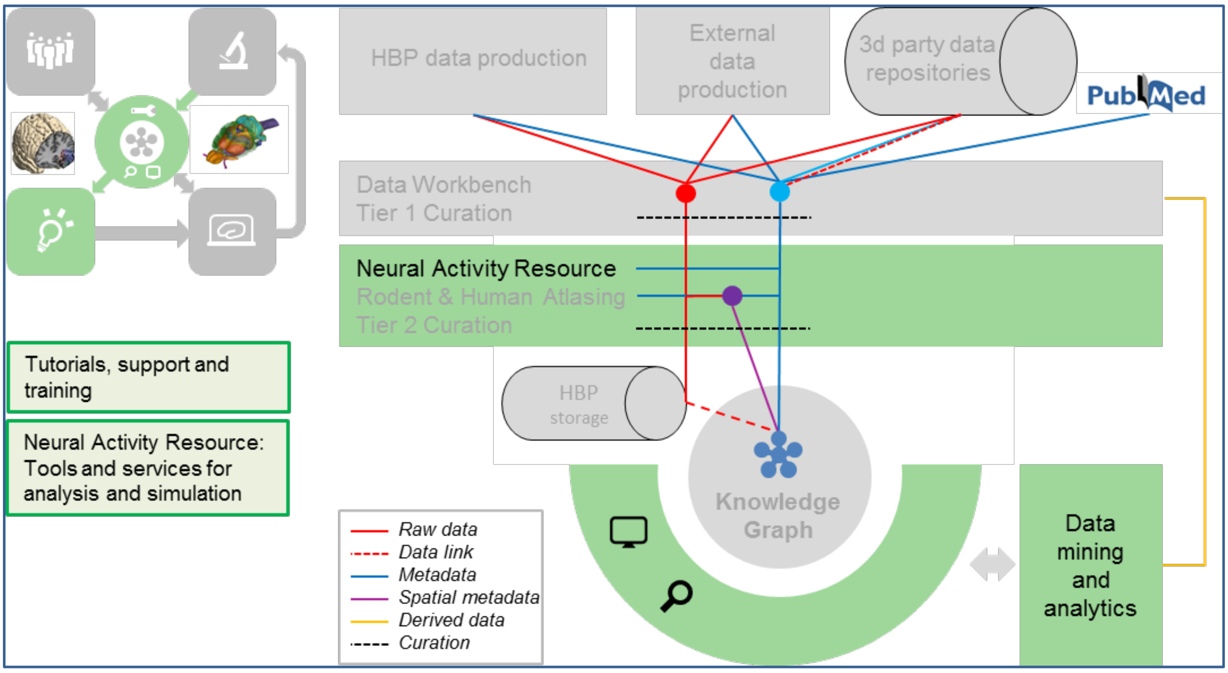

Figure 24: Schematic overview with contributions from WP5.7, highlighted in green.

8.1 Key Personnel

WP Leader: Sonja GRÜN (JUELICH)

Task Leaders: Sonja GRÜN (JUELICH)

Andrew DAVISON (CNRS)

8.2 Work Package Leader's Overview

During the RUP, the HBP had a good coverage of anatomical data and atlases. Activity data, such as spiking activity and population activity data, had been relatively neglected. However, the results of neuronal network simulations, which constitute a central aspect of the HBP (e.g., using NEST, Neuron, SpiNNaker and BrainScales), all generate activity data of this kind [spiking activities and local field potentials (LFPs)]. Indeed, one of the declared goals of the HBP is to validate network simulations by comparison to experimental data. Thus, with the NAR component introduced in the revised work plan, we will now provide the collective resource for the management of neural activity data and its metadata needed in the HBP for such validation tests of the neural network models. The NAR is complemented by development of the Electrophysiology Analysis Toolkit (Elephant), which provides the tools for analysis of such data from experiment or simulation in a unified data framework (implemented by the Neo software library). Thus, the tools developed by this Task aim to support the development of reproducible and structured data analysis workflows within the HBP, facilitating collaboration and transparent publication of research.

Key achievements in the first half of SGA1 are the 0.5 release of the Neo and 0.4 release of Elephant library. These constitute a major step towards simplifying the activity data model and increasing the compatibility between these interlinked libraries. In addition, the team started the development of optimised, parallel versions of computationally demanding analysis methods to prepare for upcoming parallelisation efforts of the library planned in

SGA2. To this end, T5.7.1, in collaboration with the Simlab Neuroscience at JUELICH, has taken one such state-of-the-art analysis method, termed ASSET (Torre *et al.*, 2016), and optimised/parallelised it under the principles of co-design: Code was optimised using both general principles, and specific knowledge of the PCP architecture by a software engineer, while the optimised code was continuously and iteratively validated on HPC machines using mock data (simulating the features of the targeted batch analysis scenarios) by a computational neuroscientist. The process achieved a significant and sufficient speed-up of the analysis, and acts as a model for the co-design process of upcoming optimisations where necessary (presented as poster at the HBP Student Conference 2017).

With regard to the NAR, we were recently able to hire one person (Mehmet SUZEN) for T5.7.2 after the delayed start of SP5. The hiring of a second person for T5.7.2 is in progress. An initial set of datasets for inclusion in the NAR has been identified, and the initial concept for the NAR was designed. One dataset of a complex electrophysiological experiment has been documented with metadata in extensive detail in order to serve as a standard example to guide NAR development for electrophysiology data on the level of spikes and local field potential data.

People involved in the WP meet regularly and got quickly started this year; discussions with other WPs within SP5, but also with other SPs are taking place regularly and help enormously to integrate our work into the rest of the platforms.

Communication within SP5 is transparent and fast; collaboration with other WPs works smoothly.

8.3 Priorities for the remainder of the phase

Priorities for the Elephant tool involve the further integration of additional functionality as well as continued work on the experimental parallelisation support for selected functions (in particular ASSET and SPADE) to prepare for integration of extended parallelisation capabilities in SGA2. Additional functionality includes, but is not limited to, LFP/LFP-spike (phase) analysis (collaboration with SP3), the integration of a novel method for spatio-temporal spike pattern detection and significance estimation, advanced spectral methods (CDP4), and the implementation of the Gaussian Process Factor Analysis. Also, this Task will further work on improving interfaces between Elephant/Neo and data obtained from the NEST simulator, beyond the corresponding I/O developed for Neo 0.5 in M12.

With regard to the NAR (T5.7.2), priority will be the integration of datasets that were identified up to M12. This includes the initial setup of the Tier 2 curation of these data by defining the corresponding metadata describing these datasets. This will be done in close interaction with the data providers. Of particular priority is the integration of the data component “Massively Parallel Electrophysiology Data”, which in the context of preparatory work for T5.7.2 has received an extensive set of metadata giving a rich, detailed account of a complex behavioural experiment involving multi-scale activity data. Coming from this detailed description, the Task will be able to extract those metadata that are of high interest for the curation in the NAR.

8.4 Milestones

See Table 7.

Table 7: Milestones for WP5.7 — Tools and Curation for Integrated Parallelised Analysis of Activity Data

MS No.	Milestone Name	Lead Partner	Task(s) involved	Month Due	Month of Completion	Comments
MS5.7.1	List of KPIs for internal and external user adoption of Products/services offered by WP5.7	P20(JUELICH)	T5.7.1 T5.7.2	M10	M10	Revised and finalised by SP Leader. Approved by Task/WP leaders. Minor adjustments of KPIs for internal SP use will be considered based on accumulated experience during the project period. A pre-verification version of this Milestone was delivered to the European Commission together with the D5.6.1 Resubmission in January 2017. The full list of KPIs is available for the HBP SGA1 consortium through the Milestone Report Repository and can also be found in Appendix 1. In this report, we include the M18 report on KPIs, with comments on adjustments of targets for SGA1

8.5 T5.7.1 – Elephant

8.5.1 *Key Personnel*

Task Leader: Sonja GRÜN (JUELICH)

Other Researcher: Michael DENKER (JUELICH)

8.5.2 *SGA1 DoA Goals*

The Task will coordinate the further development of the Elephant data analysis toolkit for the analysis of functional data from simulation and experiment. Key aspects are the integration of additional analysis modules, prototype integration of support for parallelisation of selected analysis methods, further development of the Neo and its file interfaces including NEST connectivity.

8.5.3 *Component Progress*

NOTE: This Task report includes a new Component (115-3b) which was not listed in the work plan. The new Component, Neo, closely related to Elephant (115-3), is a Python library for representing electrophysiology data in Python, together with support for reading a wide range of neurophysiology file formats. It is a dependency of Elephant, and is also used by tools in SP9 (PyNN) and SP6 (eFEL). Neo is developed by an open-source community, to which we contribute.

8.5.3.1 Component 113-1 Elephant

Completed in RUP.

8.5.3.2 Component 114-2 Elephant Code Repository

Completed in RUP.

8.5.3.3 Component 115-3 Elephant

Description of Component: Elephant is a toolbox for the analysis of electrophysiological data based on the Neo framework.

Progress on Component: Note: Elephant development has been moved from T9.1.5 to T5.7.1 during the DPIT process.

Elephant has received updated functionality through inclusion of a state-of-the-art current-source-density analysis algorithm. In addition, there has been significant progress on implementing one of the most powerful methods available today to analyse patterns of synchronous spiking activity (SPADE), and the method is planned to be made available in the upcoming release of Elephant. Efforts to add parallelisation support have been initiated in collaboration with SP7 and the Simlab Neuroscience at JUELICH by adding parallelisation to two methods to detect higher-order correlation structure in spike data (ASSET and SPADE), in part optimised for the new PCP machines at JSC. Both of these complementary methods are expected to play an important role in the analysis and validation of simulation data, and their parallelisation is required to probe the scalability of the analysis to the amount of data obtained from network simulations. The optimisation for the ASSET method has provided significant speed-ups that have been presented in poster form at the 2017 HBP student conference; these enhancements will be made available in Elephant during the second half of SGA1. Further work on Elephant included adaptation of functions to the major API changes of the Neo library version 0.5, which led to the 0.4 release of Elephant. Furthermore, JUELICH has developed the first prototype of a data analysis workflow for experimental data, encompassing the acquisition of data, annotation with metadata, and analysis, using the Collaboratory. These hands-on results were presented during the HBP Student Conference and at the German Neuroscience Meeting in March 2017.

Links:

Elephant 0.4 release: https://services.humanbrainproject.eu/software-catalog/catalog/?ctx=97d59245-adce-42c4-8b50-d22587071fcd#?ctxstate=software,Elephant.

Data analysis workflow for experimental data: https://collab.humanbrainproject.eu/#/collab/1710/nav/15298

8.5.3.4 Component 115-3b Neo

Description of Component: Neo is a package for representing electrophysiology data in Python, together with support for reading a wide range of neurophysiology file formats, including Spike2, NeuroExplorer, AlphaOmega, Axon, Blackrock, Plexon, Tdt, and support for writing to a subset of these formats plus non-proprietary formats including HDF5. The goal of Neo is to improve interoperability between Python tools for analysing, visualising and generating electrophysiology data (such as OpenElectrophy, NeuroTools, G-node, Helmholtz, PyNN) by providing a common, shared object model. In order to be as lightweight a dependency as possible, Neo is deliberately limited to representation of data, with no functions for data analysis or visualisation. Neo implements a hierarchical data model well adapted to intracellular and extracellular electrophysiology and EEG data with support for multi-electrodes (for example tetrodes). Neo's data objects build on the quantities package, which in turn builds on NumPy by adding support for physical dimensions. Thus, Neo objects behave just like normal NumPy arrays, but with additional metadata, checks for dimensional consistency and automatic unit conversion. A project with similar aims but for neuroimaging file formats is NiBabel.

Progress on Component: In M1-M12 Neo, the data model underlying the Elephant data analysis framework but also other tools integrated in the HBP Platforms, has undergone a significant API change to simplify working with the library and to reduce the overall complexity of the model. These changes have been released as a new major version 0.5 and are accompanied by a matching Elephant release 0.4 reflecting these changes. In terms of data interoperability, T5.7.1 has worked on three additional file I/O modules, which have either entered the Neo 0.5 release or are accepted in the main branch of the Neo repository: (i) a file I/O for the Blackrock format, important to work on the data delivered as Component "Massively Parallel Electrophysiology Data" and for completion of T4.5.1 (joint work with T5.7.2); (ii) a file I/O for the Neuralynx format, a major provider of neuronal recording hardware and expected to be the basis for numerous datasets provided for the NAR (T7.5.2); and, (iii) the NEST file I/O which handles the output file format of the NEST simulator.

Links:

Neo 0.5 release: https://services.humanbrainproject.eu/software-catalog/catalog/?ctx=97d59245-adce-42c4-8b50-d22587071fcd#?ctxstate=software,neo

8.6 T5.7.2 - Neural Activity Resource

8.6.1 *Key Personnel*

Task Leader: Andrew DAVISON (CNRS)

8.6.2 *SGA1 DoA Goals*

The goal is to define the process of the Tier 2 curation of activity data sets generated by the HBP, and to develop the tools, in conjunction with the data workbench developed by WP5.1, that allow users to register, search, and retrieve such datasets. In an initial prototype, this Task will develop strategies to link data to the HBP atlas.

8.6.3 *Component Progress*

8.6.3.1 Component 116-1 Neural Activity Resource

Description of Component: For integration and analysis of neural dynamics data (ECoG, fMRI, cellular, ensembles, model results, behaviour, etc.) Products produced in T5.1.1 and operationally supported by the work in T5.7.1. The development in T5.7.2 is the in-depth adaption of the resource together with the users.

Progress on Component: Mehmet SUZEN has been hired at JUELICH to start work on the NAR component. Eight activity data sets have been identified across the HBP that will compose the initial focus for integration. These datasets span different measurement modalities, different SPs, and include experimental as well as simulated data. In preparatory work for T5.7.2 and in the context of WP4.5, the annotation of one of these datasets (component Massively Parallel Electrophysiology Data) with an extensive set of metadata in the odML metadata format has been completed, giving the opportunity to nominate from this description the essential metadata to enter the Tier 2 curation carried out by the NAR. Based on the odML metadata framework, a first version of the workflow for integrating activity data has been defined.

8.6.3.2 Component 117-2 Linkage of Neural Activity Resource to Human and Mouse brain atlases

Description of Component: Linkage between Neural activity resource to Human and Mouse brain atlas.

Progress on Component: Work has not yet started on this Component due to the only recent hiring of personnel.

8.6.3.3 Component 115-3b Neo

Description of Component: See T5.7.1.

Progress on Component: T5.7.2 has contributed to the development of a Neo file I/O module for the Blackrock format in collaboration with T5.7.1. This component is required to integrate the data delivered as component "Massively Parallel Electrophysiology Data" into the NAR framework.

9. WP5.8 Management and Coordination

9.1 Key Personnel

SP Leader: Jan BJAALIE (UIO)

SP Deputy Leader: Sten GRILLNER (KI)

SP Manager: Sofia ANDERHOLM STRAND (UIO)

9.2 Work Package Leader's Overview

UIO gradually ramped-up management and general coordination of SP5 since M7 to full capacity at M11/M12. The technology and infrastructure coordination was performed in close collaboration with the HBP Technical Coordinator, who is also leader of WP5.5 of SP5.

KI has continued the dissemination activities and the development of KnowledgeSpace.

A major part of the coordination effort up to January 2017 was coupled to the preparation of D5.6.1 and the D5.6.1 Resubmission.

WP5.8 has facilitated internal communication and collaboration, including co-developments of products, as well as collaborations with other SPs. In addition to regular meetings (in-person or videoconferencing) and email communications, the Collaboratory has been used extensively.

9.3 Priorities for the remainder of the phase

A priority in the remainder of the phase will be to further strengthen the capacity for SP5 technical and infrastructure coordination, in a continuing close interaction with the HBP Technical Coordinator.

9.4 Milestones

See Table 8.

Table 8: Milestones for WP5.8 — Management and Coordination

MS No.	Milestone Name	Lead Partner	Task(s) involved	Month Due	Month of Completion	Comments
MS5.8.1	List of KPIs for performance evaluation of coordination, outreach and adoption of products/services offered by WP5.8	P81(UIO)	T5.8.1 T5.8.2 T5.8.3	M10	M10	Revised and finalised by SP Leader. Approved by Task/WP leaders. Minor adjustments of KPIs for internal SP use will be considered based on accumulated experience during the project period. A pre-verification version of this Milestone was delivered to the European Commission together with the D5.6.1 Resubmission in January 2017. The full list of KPIs is available for the HBP SGA1 consortium through the Milestone Report Repository and can also be found in Appendix 1. In this report, we include the M18 report on KPIs, with comments on adjustments of targets for SGA1

9.5 T5.8.1 - Subproject Management and Technical Coordination

9.5.1 *Key Personnel*

SP Leader: Jan BJAALIE (UIO)

SP Manager: Sofia ANDERHOLM STRAND (UIO)

9.5.2 *SGA1 DoA Goals*

Facilitating internal communication, resource allocation and use, performance and risk management, internal review, quality control and reporting.

Coordinating the technology and infrastructure development.

9.5.3 *Progress on Task*

UIO has gradually ramped-up management and general coordination of SP5 since M7 to full capacity at M11/M12. The technology and infrastructure coordination has been performed in close collaboration with the HBP Technical Coordinator, who is also leader of WP5.5 of SP5. A major part of the coordination effort up to January 2017 was coupled to the preparation of the D5.6.1 Resubmission. In addition to regular meetings (in-person or videoconferencing) and email communications, the Collaboratory has been used extensively.

9.6 T5.8.2 - Neuroinformatics Community Building

9.6.1 *Key Personnel*

Task Leader: Sten GRILLNER (KI)

Other Researcher: Juan PEREZ-FERNANDEZ (KI)

9.6.2 *SGA1 DoA Goals*

Supporting community activities, including:

- Acting as liaison between HBP and INCF;
- Organisation of workshops to promote the broad usage of Neuroinformatics in the community;
- Coordinating communications towards participation of different neuroscience communities in Europe;
- Assisting the SIB chair for communications worldwide.

9.6.3 *Progress on Task*

Interactions with INCF have continued and have had a particular focus on the KnowledgeSpace development. Extensive interaction has taken place with the different brain initiatives in meetings and collaboration with different stakeholders on a global scale and in interaction with the SIB chair:

- Sten Grillner organised a Keystone conference on the State of the Brain in Alpbach, 22-26 May 2016 with participation among others of the EU director Thomas SKORDAS and the head of NINDS Walter KOROSHETZ.
- SG took part in the Rockefeller meeting, New York, on Coordinating Brain Projects across the Globe on 19 September 2016 and provided the wrap up concluding the meeting.
- SG took part in the "International Brain Initiative meeting at UN, New York on 27 February 2017.

- SG presented at a Meeting on Janelia Farm, US, Motor Control Circuits – structure, function and behaviour, 9-12 May 2016.
- SG presented at third INT conference Marseille, 15-17 September 2016.
- Presentation at a Meeting on Janelia Farm, US, Action Selection across the Animal Kingdom, 26-28 September 2016.
- SG discussed with The Swedish Minister of Research regarding Neuroinformatics and HBP on 15 February 2017.
- SG participated in biannual meeting of FENS and in this context discussed the HBP extensively, SG participated in the Society for Neuroscience in November 2016 where Knowledge Space was released in interaction with INCF.

Workshops for using the NIP will be organised when the platforms are released.

9.7 T5.8.3 - Development and Maintenance of KnowledgeSpace

9.7.1 *Key Personnel*

Task Leader: Sten GRILLNER (KI)

Other Researcher: Mathew ABRAMS (KI)

9.7.2 *SGA1 DoA Goals*

KnowledgeSpace is a community encyclopaedia and consists of interactive interface and data-driven taxonomies, production ontology services, data type.

9.7.3 *Component Progress*

9.7.3.1 Component 118-1 Knowledge Space

Completed in RUP.

9.7.3.2 Component 119-2 Knowledge Space

Completed in RUP.

9.7.3.3 Component 120-3 Cell type ontologies

Description of Component: per brain area: hippocampus, cerebellum, basal ganglia, cortex - including key molecular markers and property-based cell ontology for each region.

Progress on Component: This activity will be prioritised in the remainder of the phase in collaboration with WP5.1.

9.7.3.4 Component 121-4 Integrated data sources

Description of Component: Integration of additional federated data sources to support the modelling Use Cases; Search results for current federated data sources tuned to support modelling Use Cases; Subscribe/notify for concepts - receive notifications when new data becomes available in the searchable dataspace.

Progress on Component: We have reached a major Milestone in our efforts to develop a community-driven wiki that provides an encyclopaedic view of the latest data, models, and literature of brain organisation with the release of KnowledgeSpace 1.0 at the meeting of Society for Neuroscience in San Diego. This release serves as a proof of concept for our strategy to link concepts about the brain to data, models, and literature. During this phase, we have focused on linking concepts to data, models, and literature from existing community resources representing a limited range of levels of the mouse brain organisation.

10. Co-Design Project Number and Name

SP5 is not the leader of a CDP but contributes to several. We refer to the reports from the SPs leading the CDPs.

11. Publications (to date, open access marked with bold)

Antón Sánchez L. Statistical and optimization methods for spatial data analysis applied to neuroscience, dissertation at the UPM (P68**)** http://oa.upm.es/47383/ Related to T1.4.1: Synapses spatial location on the dendritic shaft

Anton-Sanchez L, Bielza C, Larrañaga P, DeFelipe J. (2016). "Wiring economy of pyramidal cells in the juvenile rat somatosensory cortex", ***PLoS ONE***;11(11):e0165915. T5.3.4.

Cazemier JL, Clasca F, Tiesinga PH. (2016). Connectomic Analysis of Brain Networks: Novel Techniques and Future Directions. ***Frontiers in Neuroanatomy*** 10;110, doi:10.3389/fnana.2016.00110. T5.2.5

Dempsey B, Le S, Turner A, Bokiniec P, Ramadas R, Bjaalie JG, Menuet C, Neve R, Allen AM, Goodchild AK, McMullan S. (2017). Mapping and Analysis of the Connectome of Sympathetic Premotor Neurons in the Rostral Ventrolateral Medulla of the Rat Using a Volumetric Brain Atlas. ***Frontiers in Neural Circuits***;11:9. doi: 10.3389/fncir.2017.00009. T5.2.2

Diekhoff-Krebs SN, Sarfeld AS, Rehme AK, Eickhoff SB, Fibk GR, Grefkes C (2017). Interindividual differences in motor network connectivity and behavioral response to iTBS in stroke patients. *NeuroImage: Clinical* https://doi.org/10.1016/j.nicl.2017.06.006

Du K, Nonlinear synaptic integration on dendrites of striatal medium spiny neuron – a computational study, Thesis of the Karolinska institutet 2017, 1-60 . (ISBN 978-91-7676-527-2, https://openarchive.ki.se/xmlui/handle/10616/45395). T5.1.2.4, T5.8.2, T6.1.5, T6.2.4

Eickhoff SB, Constable RT, Yeo BT. Topographic organization of the cerebral cortex and brain cartography. (2017) *Neuroimage;* pii: S1053-8119(17)30122-2. doi: 10.1016/j.neuroimage.2017.02.018. [Epub ahead of print]. T5.4.4

Genon S, Reid A, Li H, Fan L, Müller VI, Cieslik EC, Hoffstaedter F, Langner R, Grefkes C, Laird AR, Fox PT, Jiang T, Amunts K, Eickhoff SB. (2017). The heterogeneity of the left dorsal premotor cortex evidenced by multimodal connectivity-based parcellation and functional characterization. *Neuroimage;* pii: S1053-8119(17)30145-3. doi: 10.1016/j.neuroimage.2017.02.034. [Epub ahead of print]. T5.4.4

Grillner S, Robertson B. (2016). The Basal Ganglia Over 500 Million Years. ***Current Biology***;26(20):1088-1100. T5.1.2.4, T5.8.3, T6.1.5, T6.2.4

Grillner S, Ip N, Koch C, Koroshetz W, Okano H, Polachek M, Poo MM, Sejnowski TJ. (2016). Worldwide initiatives to advance brain research. *Nature Neuroscience*;19(9):1118-22. T5.1.2.4, T5.8.2, T6.1.5, T6.2.4

Grillner S. Commentary on F1000 article by Vogelstein entitled "'Grand challenges for global brain sciences" ***F1000Research*** **2016**, 5:2873 (doi: 10.12688/f1000research.10025.1). T5.1.2.4, T5.8.2, T6.1.5, T6.2.4

Hinne M, Meijers A, bakker R, Tiesinga PH, Mørup M, van Gerven MA. (2017). The missing link: Predicting connectomes from noisy and partially observed tract tracing data. ***PLoS Computational Biology***;13:e1005374, doi:10.1371/journal.pcbi.1005374 (2017). T5.2.5

Konyushkova K, Sznitman R, Fua P. (2017). "Learning Active Learning from Real and Synthetic Data". ***arXiv*:**1703.03365v2, 2017. T5.6.3

Lindahl M, Hellgren Kotaleski J. (2016). Untangling Basal Ganglia Network Dynamics and Function: Role of Dopamine Depletion and Inhibition Investigated in a Spiking Network Model. ***eNeuro*;3(6**): pii: *eNeuro*.0156-16.2016. T5.1.2.4, T5.8.2, T6.1.5, T6.2.4

Pavlovic M, Zacharatou ET, Sidlauskas D, Heinis T, Ailamaki A. (2016). "Space odyssey: efficient exploration of scientific data." In Proceedings of the Third International Workshop on Exploratory Search in Databases and the Web (pp. 12-18). **ACM.** [https://dl.acm.org/citation.cfm?id=2948677]. T5.2.7

Pläschke RN, Cieslik EC, Müller VI, Hoffstaedter F, Plachti A, Varikuti DP, Goosses M, Latz A, Caspers S4, Jockwitz C, Moebus S, Gruber O, Eickhoff CR, Reetz K, Heller J, Südmeyer M, Mathys C, Caspers J, Grefkes C, Kalenscher T, Langner R, Eickhoff SB (2017). On the integrity of functional brain networks in schizophrenia, Parkinson's disease, and advanced age: Evidence from connectivity-based single-subject classification. *Hum Brain Mapp*. DOI: 10.1002/hbm.23763

Rostami V, Ito J, Denker M, Grün S (2017) [Re] Spike synchronization and rate modulation differentially involved in motor cortical function. ***Re:Science*,** https://github.com/ReScience-Archives/Rostami-Ito-Denker-Gruen-2017/blob/master/article/Rostami-Ito-Denker-Gruen-2017.pdf

Schubert N, Axer M, Schober M, Huynh AM, Huysegoms M, Palomero-Gallagher N, Bjaalie JG, Leergaard TB, Kirlangic ME, Amunts K, Zilles K. (2016) 3D Reconstructed Cyto-, Muscarinic M2 Receptor, and Fiber Architecture of the Rat Brain Registered to the Waxholm Space Atlas. ***Frontiers in Neuroanatomy***;10:51. doi: 10.3389/fnana.2016.00051. T5.2.1

Torre E, Canova C, Denker M, Gerstein G, Helias M, Grün S (2016). ASSET: Analysis of sequences of synchronous events in massively parallel spike trains. ***PLOS Computational Biology*** 12, e1004939

12. Dissemination Plan

The objective of SP5 is to support the activities of other SPs by developing and maintaining infrastructure and services for organising, managing, visualising, analysing, and openly sharing multi-level and multi-modal research data and computational models.

Therefore, SP5 acts as a liaison between the different SPs to improve the overall impact of the HBP by making HBP data discoverable and accessible. Within SP5, WP5.8 is responsible for community building, project management and coordination with help from other WPs. In addition to that, SP5 continuously creates awareness of the project among broader audience and scientific communities outside HBP.

12.1 Plan for engaging the Community

12.1.1 *Target audience*

Key stakeholders for SP5 have been identified and are listed in Table 9.

Table 9: List of target audiences and relevant stakeholders

Target audience	Why them?	What is in it for them?
HBP partners	The HBP partners will use the tools and services for organising, managing, integrating, and analysing the heterogeneous research data that they are either providing or making use of.	Support for findable, accessible, interoperable, reusable (FAIR) data. Workflows for analysis and extraction of features relevant for neuroscience research and computational modelling.

EU/world scientists	Scientists outside HBP can use HBP services to search and retrieve the openly shared data and use the HBP tools and services for data analysis.	Discovery and hypothesis driven research using HBP data, tools, and services. Workflows for analysis and extraction of features relevant for neuroscience research and computational modelling. Atlases of integrated research data.
Graduate students	Involvement of young scientists into future development of EU science.	Training in neuroinformatics methods involving digital brain atlasing resources, automated quantitative analysis, and data mining approaches utilising shared neuroscience data. Network for career opportunities.
Industry	Research divisions of companies can use HBP services to search and retrieve the openly shared data and use the HBP tools and services for data analysis.	Discovery and hypothesis driven research using HBP data, tools, and services. Workflows for analysis and extraction of features relevant for neuroscience research and computational modelling. Atlases of integrated research data.

12.1.2 *Outreach strategy*

12.1.2.1 Scientific publications

The scientific community will be informed throughout the project about SP5's scientific achievements primarily through scientific publications, peer-reviewed journals, conference abstracts, proceedings, talks and posters. SP5 strongly promotes an open access publication policy to reach a larger amount of scientists worldwide in line with Horizon 2020 objectives. Due to the nature of SP5 we promote cross-SP publications based on the tools of our platform. All publications published by SP5 will include acknowledgement of the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 720270 (HBP SGA1).

12.1.2.2 Workshops

Workshops and similar events will be used by SP5 to increase awareness about data visualisation and neuroinformatic tools. Such events will engage both specialists and students thus involving both dissemination and exchange of knowledge as well as generation of new ideas, building platforms for future collaborations. The workshops will also serve as a platform for young researchers involved in SP5 to gain more "soft" skills in communication and network building, improving their future career perspectives in line with Horizon 2020 policies.

12.1.2.3 Training of young scientists.

The new crop of European scientists, the young master and PhD students will be offered online tools with tutorials, hands-on courses and workshops. At SP5 meetings, conferences, workshops and summer schools, students and young scientists will benefit from a direct interaction with renowned scientists. This embodies an excellent opportunity for young scientists, who will be appointed when funding is obtained, to start building an international scientific network, acquire new skills and to develop their scientific careers. With the opportunity to attend workshops, summer schools and SP5 open meetings, we are able to reach the young scientific community and safeguard the influx of eager and knowledgeable future leading expert scientists to the field and sustain state of the art and innovative research on the brain in Europe. Many of those events are organised together with HBP educational programme and national training offers.

12.1.2.4 Presentations at conferences, open days and public lectures.

SP5's tools and results will be presented at international and national conferences, such as the Society for Neuroscience Annual Meeting or Nordic Neuroscience meeting, amongst others. The SP5 leadership, participants and young researchers will also attend and present the vision of the project at relevant professional events at their partner institutions. This will allow us to reach the relevant stakeholders in a convenient way.

SP5 scientists will actively participate in HBP open days at HBP Summits and other public opportunities to disseminate the information about SP5 research and software to a broader audience. This will further increase the impact of the project.

12.1.2.5 Social media and public website strategy

We plan to have ongoing releases of SP5 outputs to the wider community starting in SGA1. Initial releases of the reference atlases have already been made through the Explore the Brain section of https://www.humanbrainproject.eu. These updates are expected to continue throughout SGA1. There are several other types of releases which will be carried out by SP5, many of them on behalf of other areas of HBP through the efforts of the Tier 1, Tier 2 and Tier 3 curation process. It is assumed that all data releases are also reflected in the KG and are assigned a DOI. Elephant tool will get a deducated website at the start of SGA2 to host all relevant information, tutorials and updates.

Reference atlas update

1) Public release to https://www.humanbrainproject.eu/en/explore-the-brain/atlases/.
2) Public website news item with notes describing changes and link to 1.
3) HBP Twitter (@HumanBrainProj), HBP Collaboratory Twitter (@HBPCollaborate) and HBP Facebook postings linking to the news item.

Dataset releases

Once the KG search has had its initial public release, data releases should also be publicised in the same way as above. Each release will be given a dataset overview card which will be reachable through the KG search and this will serve as the public home for the dataset in HBP. Dataset releases should happen at the completion of Tier 1 curation or at the end of the embargo period, whichever takes place later. Embargoed datasets should not be publicised until the data becomes available, even though they will be visible in the KG search.

1) A DOI will be assigned and linked to the public dataset overview card as part of the KG registration process.
2) HBP Twitter (@HumanBrainProj), HBP Collaboratory Twitter (@HBPCollaborate) and HBP Facebook postings linking to the DOI.
3) If scientists want to time the release of their data for maximum impact, they should work with the HBP public website team to write a custom ***news item*** (https://www.humanbrainproject.eu) highlighting the value and community relevance of the dataset in question. If scientists do not want to provide such content, dataset summaries will be generated from Tier 1 metadata as described below.
4) Individual datasets or clusters of related datasets should be provided with a https://www.humanbrainproject.eu news item describing the value of the dataset and containing thumbnails for the data.

Tools releases

Major tools releases will be announced through HBP Twitter (@HumanBrainProj), HBP Collaboratory Twitter (@HBPCollaborate) or HBP Facebook postings, linking also to relevant tutorials.

Other: Press releases, Interviews

Press releases, as well as radio interviews and social media groups will be planned to promote the project and to communicate its major achievements to the general public. The participants have experience in giving media interviews.

12.2 Progress in Outreach

Already in the first year of SGA1, SP5 members have been successful in publishing HBP-related and HBP-sponsored research in many journals, including high-ranking ones such as Nature Neuroscience. In addition, several outreach events and meetings were organised together with other SPs (table 2) and SP5 made important contributions to several HBP-based public events, such as the HBP open days at the HBP Summits in Florence and in Glasgow, in which, for example, the virtual reality and 3D atlas demonstration was displayed.

Members of SP5 give lectures about their HBP research in public/open and professional fora several times per year (e.g. at seminars, schools, symposia) for neuroscientists and IT specialists.

SP5 works in a close collaboration with SP1, SP2 and SP3, processing data produced in these SPs and making it accessible to others through SP5, SP6 and SP7. The HBP consortium is aware of the tools being developed by SP5 and we expect to see a dramatic increase in user numbers when the platform is released. Different meetings with HBP partners are being held regularly for the continuance of solid co-operations into SGA2 and SGA3 periods.

In addition to active engagement of SP5 in various events where young scientists gain advanced knowledge about HBP, SP5 participants train 16 PhD students (one graduated), nine junior scientists (postdocs) and a number of master students, thus ensuring the continuous knowledge transfer to younger generation of European researchers. Importantly, SP5 pays attention to gender balance and 45% of our PhD and postdocs (junior scientists) are female.

12.2.1 *List of main scientific events with active contribution from SP5 (to date)*

Contributing PIs are indicated in **bold**.

“Computational Intelligence and Neuroscience”: Institute of Engineering Schools of Spain, Madrid, 2016. Components ‘Synapse spatial location on the dendritic shaft model’, ‘GABAergic interneuron classifier’, ‘Spine morphology clustering’, and ‘Single cell arborisation model’. T1.4.1 and T5.3.4. **Larrañaga P**

INCF Neuroinformatics 2016: "Web-based neuron morphology viewer as an aid to develop new standards for neuron morphology file formats", Reading UK, 3-4 September 2016. T5.2.5. Bakker R, **Tiesinga P**

PINC conference: “Building Brains”, 10 May 2016. **Tiesinga P**

RDA workshop: “Neuroscience Data”, 19 April 2016. **Tiesinga P**

ODEN (Open Data Ecosystem) meeting: “No title” panel discussion on open data issues in HBP and beyond, 25 July 2016. T5.2.5. **Tiesinga P**

Third International Workshop on Exploratory Search in Databases and the Web (co-located with SIGMOD/PODS 2016): "Space odyssey: efficient exploration of scientific data", San Francisco, California, USA, 1 July 2016. Sidlauskas D, **Ailamaki A**

Keystone conference on the "State of the Brain", Alpbach, 22-26 May 2016 with participation among others of the EU director Thomas Skordas and the head of NINDS Walter Koroshetz. **Grillner S**

Rockefeller meeting: "Coordinating Brain Projects across the Globe", New York, 19 September 2016. **Grillner S**

"International Brain Initiative meeting" at UN, New York, 27 February 2017. **Grillner S**

"Motor Control Circuits – structure, function and behaviour": meeting on Janelia Farm, US, 9-12 May 2016. **Grillner S**

Third INT conference, Marseille, 15-17 September 2016. **Grillner S**

"Action Selection across the Animal Kingdom": meeting on Janelia Farm, US, 26-28 September 2016. **Grillner S**

Third Annual BRAIN Initiative Investigators Meeting, National Institutes of Health, Bethesda, 12-14 December 2016. **Bjaalie JG**

Understanding the brain through neuroinformatics: from experiments to modelling and theory. Public meeting on the Human Brain Project, Embassy of Switzerland, Washington DC. 13 December 2016. **Bjaalie JG**

Discussion with the Swedish Minister of Research regarding Neuroinformatics and HBP, 15 February 2017. iEEG

Sten Grillner participated in the FENS biannual meeting and this context discussed HBP extensively and in the Society for Neuroscience in November 2016 where Knowledge Space was released in interaction with INCF. **Grillner S**

"Machine learning for image analysis: establishing pipelines for high throughput feature extraction" INCF Nodes Workshop, 29-30 Oslo 2017. **Kreshuk A**

"Tools for teaching neuroanatomy" INCF Nodes Workshop, 29-30 Oslo 2017, **Leergaard T**

"Integrating experimental neuroscience data in rodent brain reference atlases" INCF Nodes Workshop, 29-30 Oslo 2017, **Leergaard T**

"Towards a multi-level human brain atlas: A challenge in bridging scales and addressing variability" INCF Nodes Workshop, 29-30 Oslo 2017, **Dickscheid T**

"Infrastructure for digital brain atlasing: storing, organizing, visualizing and analyzing large quantities of heterogeneous neuroscience research data" INCF Nodes Workshop, 29-30 Oslo 2017, **Bjaalie JG**

"Building and using rodent brain reference atlases for integration of experimental data" Understanding the Brain: Neurobiology for non-specialists, HBP workshop, 3-5 July 2017 Innsbruck, **Leergaard T**

"Integrating experimental rodent brain data through digital brain atlasing" Department of Molecular Medicine retreat, UiO, Oslo 27 September 2017, **Bjerke I**

"Neuroscience Data Integration: Big Data Challenges in Brain Atlasing", XLDB, Clermont-Ferrand, 10-12 October 2017, **Bjaalie JG**

"The Human Brain Project: a Platform for Big-Data Neuroscience", WSMD2017Workshop on Schizophrenia and Other Mental Disorders, Pisa, 15-16 June 2017, **Bjaalie JG**

"The HBP Neuroinformatics Platform", Understanding the Brain: Neurobiology for non-specialists, HBP workshop, Innsbruck 3-5 July 2017, **Dickscheid T**

“The HBP Collaboratory”, HBP Young Researchers Event 2017, Geneva 12-13 September 2017, **Muller J**

“Towards reproducible analysis workflows and the role of the Neuroinformatics Platform”, HBP Young Researchers Event 2017, Geneva 12-13 September 2017, **Grün S**

“Neuroscience data analysis”, CNS 2017 Antwerp: Tutorials; Antwerp, 15 July 2017; **Denker M**

“Neuroinformatics resources” CNS 2017 Antwerp: Tutorials; Antwerp, 15 July 2017; **Davison A**

“Virtual reality demonstration of the Waxholm Space atlas of the rat brain” at the Nordic Neuroscience meeting. Stockholm, 7-9 June; Sweden. **Øvsthus M**, **Andersson K**, **Bjerke I**

“Embedding the Elephant data analysis framework into a collaborative environment” Poster presentation at INCF Neuroinformatics 2017, Kuala Lumpur, 20- 21 August 2017; **Denker M**

“Designing collaborative workflows for reproducible data analysis in electrophysiology based on the Elephant analysis framework” Poster presentation at Japan Neuroscience 2017 Meeting , 20-23 July 2017; **Denker M**

“Navigating the rodent brain: Best practice recommendations for determining and documenting spatial location for neuroscience data” Poster presentation at the INCF Nodes Workshop, 29-30 Oslo 2017, **Bjerke I**

“Data integration through digital brain atlasing: semiautomatic spatial registration of serial histological images to rodent brain 3D reference atlases” Poster presentation at the INCF Nodes Workshop, 29-30 Oslo 2017, **Puchades M**

“Teaching and researching complex biomedical structures in virtual reality: Stepping inside the Waxholm Space rat brain atlas” Poster presentation at the INCF Nodes Workshop, 29-30 Oslo 2017, **Øvsthus M**

“Virtual reality demonstration of the Waxholm Space atlas of the rat brain” demo presentation at the INCF Nodes Workshop, 29-30 Oslo 2017, **Øvsthus M**, **Wennberg A**

“Image analysis with machine learning tool ilastik” demo presentation at the INCF Nodes Workshop, 29-30 Oslo 2017; **Yates S**, **Kreshuk A**

“Navigating the rodent brain: Best practice recommendations for determining and documenting spatial location for neuroscience data” Poster presentation at the Nordic Neuroscience meeting. Stockholm, 7-9 June; **Bjerke I**

“Data integration through digital brain atlasing: semiautomatic spatial registration of serial histological images to rodent brain 3D reference atlases” Poster presentation at the Nordic Neuroscience meeting. Stockholm, 7-9 June; **Øvsthus M**

“Tackling the normalization of 2D rodent histology sections in a 3D coordinate space” Poster presentation at the Neuroinformatics 2017 Annual meeting of the INCF, Kuala Lumpur, 20-21 August 2017; **Coello C**

“Navigating the rodent brain: Best practice recommendations for determining and documenting spatial location for neuroscience data” Poster presentation at Department of Molecular Medicine retreat, UiO, Oslo 27 September 2017, **Bjerke I**

“Data integration through digital brain atlasing: semiautomatic spatial registration of serial histological images to rodent brain 3D reference atlases” Poster presentation at Department of Molecular Medicine retreat, UiO, Oslo 27 September 2017, **Øvsthus M**

“Teaching and researching complex biomedical structures in virtual reality: Stepping inside the Waxholm Space rat brain atlas” Poster presentation at Department of Molecular Medicine retreat, UiO, Oslo 27 September 2017, **Øvsthus M**

"Workflow for automated quantification and spatial analysis of labelling in microscopic rodent brain sections" Poster presentation at Department of Molecular Medicine retreat, UiO, Oslo 27 September 2017, **Yates S**.

"Virtual reality demonstration of the Waxholm Space atlas of the rat brain" demo presentation at the Department of Molecular Medicine retreat, UiO, Oslo 27 September 2017, **Øvsthus M, Kleven H, Hagen C**

"The flow of data through the Neuroinformatics Platform: data entry, storage, curation, analyses" Poster presentation at the SP5 Booth in the Science Market of the HBP Summit Open Day,Glasgow 17 October 2017, **SP5 teams at UIO, EPFL, JUELICH**

"Anatomical reference atlases for integration of experimental image data from rodent and human brains" Poster presentation at the SP5 Booth in the Science Market of the HBP Summit Open Day,Glasgow 17 October 2017, **SP5 teams at UIO, EPFL, JUELICH**

"From brain images to numbers in atlas space" Poster presentation at the SP5 Booth in the Science Market of the HBP Summit Open Day,Glasgow 17 October 2017, **SP5 teams at UIO, Heidelberg**

"From images to numbers using ilastik" Poster presentation at the SP5 Booth in the Science Market of the HBP Summit Open Day,Glasgow 17 October 2017, **Yates S, Kutra D, Beuttenmueller F**

"The HBP data workbench" SP5 Booth in the Science Market of the HBP Summit Open Day,Glasgow 17 October 2017, **Andersson K**

"HBP atlases and image registration tools" SP5 Booth in the Science Market of the HBP Summit Open Day,Glasgow 17 October 2017, **Leprince Y, Chervakov P, Gui X, Bhattacharya H, Puchades M, Bjerke I**

"Virtual reality demonstration of the Waxholm Space atlas of the rat brain" Demo at the SP5 Booth in the Science Market of the HBP Summit Open Day,Glasgow 17 October 2017, **Hagen C, Kleven H, Øvsthus M**

"From images to numbers using ilastik" SP5 Booth in the Science Market of the HBP Summit Open Day, Glasgow 17 October 2017, **Yates S, Kutra D, Beuttenmueller F**

"The HBP Collaboratory" SP5 Booth in the Science Market of the HBP Summit Open Day, Glasgow 17 October 2017, **Francani A, Santarsiero A, Hencz A, Muller J**

"Activity analysis using Elephant" SP5 Booth in the Science Market of the HBP Summit Open Day, Glasgow 17 October 2017, **Denker M, Yegenoglu A, Gutzen R, von Papen M, Suezen M**

13. Education

See Table 10.

Table 10: List of educational events hosted/co-hosted (*) or attended by SP5. The events listed below are relevant examples and not a complete list of events.

Name of event	Type	Location, Date
Budapest HBP Young Researcher Education Event	Conference	Budapest,12 April 2016
Geneva, HBP Young Researcher Event	Conference	Geneva, 12-13 September 2017
Department of Molecular Medicine (UiO) retreat	Workshop (*)	Oslo, 27 September 2017

HBP Summit Open Day in Glasgow	Public demonstrations	Glasgow, 17 October 2017
HBP Summit Open Day in Florence	Public demonstrations	Florence, 12 October 2016
Understanding the Brain: Neurobiology for non-specialists	Workshop	Innsbruck, 3-5 July 2017
Advanced Neural Data Analysis Course 2017	School (*)	Jülich, 26 March – 8 April 2017
Neuroscience Data Analysis CNS 2017	Workshop	Antwerpen, 15-20 July 2017
HBP Code Jam 2017	Workshop (*)	Lausanne, 13-15 September 2017
DATA Summit	Workshop	A Coruña, 19 June 2017
XXIV Jornadas de Classificacao e Análise de Dados	Workshop	Porto, 20-22 April 2017
XLDB	Workshop	Clermont-Ferrand, 10-12 October 2017
DEXA	Workshop	Lyon, 28-31 August 2017
INCF Nodes Workshop	Workshop (*)	Oslo, 29-30 May 2017
WSMD2017 Workshop on Schizophrenia and Other Mental Disorders	Workshop	Pisa, 15- 16 June 2017
Nordic Neuroscience Meeting	Symposia (*)	Stockholm 7-9 June 2017
INCF Neuroinformatics 2017 Annual Meeting	Conference	Kuala Lumpur, 20-21 August 2017
Japan Neuroscience Meeting 2017	Conference	Japan, 20-23 July, 2017

14. Ethics

The SP5 Ethics Rapporteur has participated in Ethics Management activities organised by SP12. In SGA1 this includes participation in video conferences and meetings for Ethics Rapporteurs, and revision of the SP5 short statement on ethical issues at M6 and M12, in collaboration with a representative of the Ethics Advisory Board and the SP5 Leader. At M12, the ethics policy for SP5 has been aligned with the Data Governance Policy developed by the Data Governance working group and prepared for implementation on the SP5 data workbench. Further details are provided in the Data Policy Manual.

15. Innovation

SP5 contributes short term to innovation with a series of releases of software and services. In the present phase of the project, the dissemination level is HBP internal, but with extensive exposure to external groups who can access and test tools. With a higher maturity level of the analytical tools and services, and following the release of large amounts of findable, accessible, interoperable, and reusable research data that can be used in combination with these services, SP5 will contribute to creating new opportunities for discoveries in basic and translational research. The broad sharing of integrated research data combined with access to tools and services represents a benefits for many sectors of society. One example is potential use of data and workflows to extract information relevant for drug developments.

16. Open Research Data

At the time of publication, several of the data-producing SPs will have made data available openly. The responsibility for SP5, in collaboration with SP7, will be to prepare the HBP storage, delivering access to data. The revised plans for HBP storage are presented in the D5.6.1 Resubmission and accompanying documents. The Data Policy Manual outlines requirements and routines related to sharing and use of data.

Appendix 1: List of KPIs for SP5 in SGA1 (MS5.1.1, MS5.2.1, MS5.3.1, MS5.4.1, MS5.5.1, MS5.6.1, MS5.7.1, MS5.8.1)

KPIs (see Table 11) will be collected at Month 18 and Month 24. All target values are cumulative and apply to SGA1 only, unless the ramp-up phase (RUP) is specifically mentioned.

Table 11: List of key performance indicators

#	KPI title/description	Target by M18	Reached by M18	Target by M24	Comments on M18 values
Data Development KPIs (Data integration, atlas building, atlas updates, etc.)					
1	Number of updated or customised versions of Allen mouse brain atlas tailored for analytical purposes, available in HBP rodent atlas list	2	3	3	Versions of Allen mouse brain atlas (v3) with reduced numbers of structures
2	Number of added or corrected structure delineations in the Waxholm Space rat brain reference atlas	20	32	40	Delineations of structures of the auditory system completed. Delineations of thalamic subregions are in preparation. The updated atlas will be shared by the end of SGA1
3	Percentage of datasets from RUP curated with metadata in KnowledgeGraph, available through KG Search API	30%	100%[a]	100%	*All RUP data made available to SP5 before M18 have been curated*
4	Percentage of SGA1 datasets shared among HBP partners and curated with metadata in KnowledgeGraph	40%	100%[a]	60%	*All SGA1 data made available to SP5 before M18 have been curated. With the new HBP storage and operative curation pipeline, a rapid increase in data flow is planned for the remainder of SGA1. The curation team is ramped up and prepared for this*
5	Percentage of curated datasets (RUP and SGA1) uploaded to HBP storage	90%	100%[b]	95%	
6	Percentage of RUP and SGA1 datasets anchored to atlas / curated with spatial metadata in KnowledgeGraph	40%	100%[a]	60%	*All SGA1 data made available to SP5 before M18 have been curated. With the new HBP storage and operative curation pipeline, a rapid increase in data flow is planned for the remaining SGA1. The curation team is ramped up and prepared for this*

7	Number of datasets from external (non-HBP) data providers prepared for entry in KnowledgeGraph	10	26	20	The 26 data sets are hippocampal connectivity data relevant for component 45-4
8	Number of datasets from external data repositories made available for entry in KnowledgeGraph	4	68	8	Data have been downloaded from the Allen Brain Institute repositories using their API. The spatial registration to reference atlas of connectivity data sets (n = 59) and in situ hybridisation data sets (n = 9) have been validated. Registration of in situ hybridisation data was improved using the HBP QuickNII tool. The workflow allows downloading of large data quantities. For the remaining period of SGA1 efforts will directed towards negotiating access to download high-resolution images from other repositories such as GENSAT and the mouse brain connectome repositories at UCLA and CSH
9	Number of datasets describing key regional characteristics of (sub)cellular elements and connectivity	4	2	8	Efforts have so far been focused on establishing a complete workflow going from images to extracted features sorted in a reference atlas mask. This has now been successfully achieved for two complete data sets. In addition, the analytical pipeline has been tuned and validated to work with a variety of image data produced within HBP. With this milestone achieved, the number of analysed data sets is expected to rapidly increase during the remaining part of SGA1
10	Number of data components evaluated for predictive approaches	1	1	2	Mouse gene expression data from the Allen Brain Atlas has been evaluated for prediction of the mesoconnectome
11	Number of activity datasets with minimal metadata accessible via the Neural Activity Resource	2	0	10	Awaiting Tier 1 curation of datasets. Preparations made. M24 target considered realistic
12	Number of activity datasets with extended metadata accessible via the Neural Activity Resource	0	0	3	
13	Number of methods for data analysis and data	43	43	45	Two methods were pending in review at end of M18 awaiting minor

	generation provided by Elephant				changes, but are integrated into the main Elephant branch early M19
Platform Development KPIs (Software and service component releases)					
14	Number of tools of maturity level TRL 2-3	1	13	2	Note that the overall number of tools reported at M18 is larger than planned
15	Number of tools of maturity level TRL 4-5 which are released to users in SP5	20	11	11	See higher values than planned for TRL 6-7 below
16	Number of tools of maturity level TRL 6-7 which are released to users in HBP	3	11	9	
17	Number of tools of maturity level TRL 8-9 which are released to users outside HBP	1	5	6	
Training and Dissemination KPIs					
18	Accumulated number of participants trained on assignment of spatial metadata to rodent or human atlas data sets	20	53	40	
19	Number of SP5 organised workshops open to the general community	1	1	2	Neuroinformatics workshop, Oslo, May 2017
20	Number of conferences or events attended with presentations of SP5 achievements (talks or posters) by SP5 partners	5	10	8	
21	Number of outreach and communication activities (contributions to press releases, interviews, and popular science articles)	2	2	3	
22	Number of early adopters contacted that browsed the BigBrain dataset in the web-based 3D viewer and reported success	1	3	10	

23	Number of early adopters that have been contacted and successfully applied landmark-based registration to align high-resolution, volumetric example data	0	1	5	
24	Number of early adopters of cellular and subcellular models of human pyramidal cell morphologies	3	0	6	Target of 6 early adopters expected to be reached at M24
25	Number of early adopters of automated detection, segmentation or counting algorithms integrated into Ilastik	3	4	6	3 early adopters related to T5.6.2; 1 related to T5.6.7
			User Activity KPIs		
26	Number of accesses to source or downloads of software components (aggregated across all SP5 downloadable software components)	20	100[c]	200	*The KPI was assumed to be per week activity numbers (3 month moving average)*
27	Number of citations of NIP hosted datasets	2	n/a	4	Not measured. KPI considered to be of limited value due to delay between publication appearance and data access
28	Number of page views on NIP	12,000	9,415	18,000	Per week activity numbers (3 month moving average). View of any NIP-owned page. Measurements aggregate page views for viewers, search, Collaboratory, etc. Collaboratory: 8751 Explore the Brain: 664
29	HBP Search requests	1,000	n/a	2,000	KnowledgeGraph not yet operative
30	Number of sessions on NIP	700	1,800	1,000	Per week activity numbers (3 month moving average). Value given for Collaboratory
31	Datasets viewed	700	n/a	1,000	KnowledgeGraph not yet operative
32	Datasets downloaded	200	n/a	400	KnowledgeGraph not yet operative
33	Number of users downloading data	50	n/a	80	KnowledgeGraph not yet operative

34	Atlas searches	200	n/a	300	This KPI would not be available until search is integrated into the Atlas viewer
35	New Jupyter notebooks created	80	166	120	Per week activity numbers (3 month moving average)
36	New collabs created	50	86	60	Per week activity numbers (3 month moving average)
37	Number of Collaboratory Apps accessed	3,500	3,693	4,000	Per week activity numbers (3 month moving average). Click on a navigation item for a Collaboratory core app (wiki, Jupyter notebook, profile, dashboard, etc.)
38	Number of non-Collaboratory SP5 Apps accessed	100	n/a	300	Tracking information currently not available in the relevant Apps
39	Number of data sets processed with ilastik	20	4	40	Progress primarily of establishing the workflows. Significant increase in number of data sets expected for remaining SGA1
<u>Internal SP5 Coordination KPIs</u>					
40	Number of internal SP5 videoconferences organised by SP5 management and coordination	10	9	15	Expected to be on target at M24
41	Number of internal SP5 physical meetings organised by SP5 management and coordination	3	3	5	

[a]Since the KnowledgeGraph is not yet operational, metadata are collected and prepared for later transfer to the KG. Note that this KPI measures the performance of the curation team, but is biased by the amount of data made available.

[b]In this report, HBP storage is defined as the available storage in Julich and at UIO, as well as Collaboratory storage. Migration of data to the new HBP storage is ongoing.

[c]Target is too low to be representative. Current download rates for two NIP software projects are in access of 100 downloads per week. List of download locations to be developed before M24.

Appendix 2: Use Case Description and Specification v 2.0

Project: Ilastik through HBP infrastructure

Partners involved: SP5, SP7

Institutions: UHEI, JUELICH, UIO

Annotated Use Case Model Diagrams

See Figure 25.

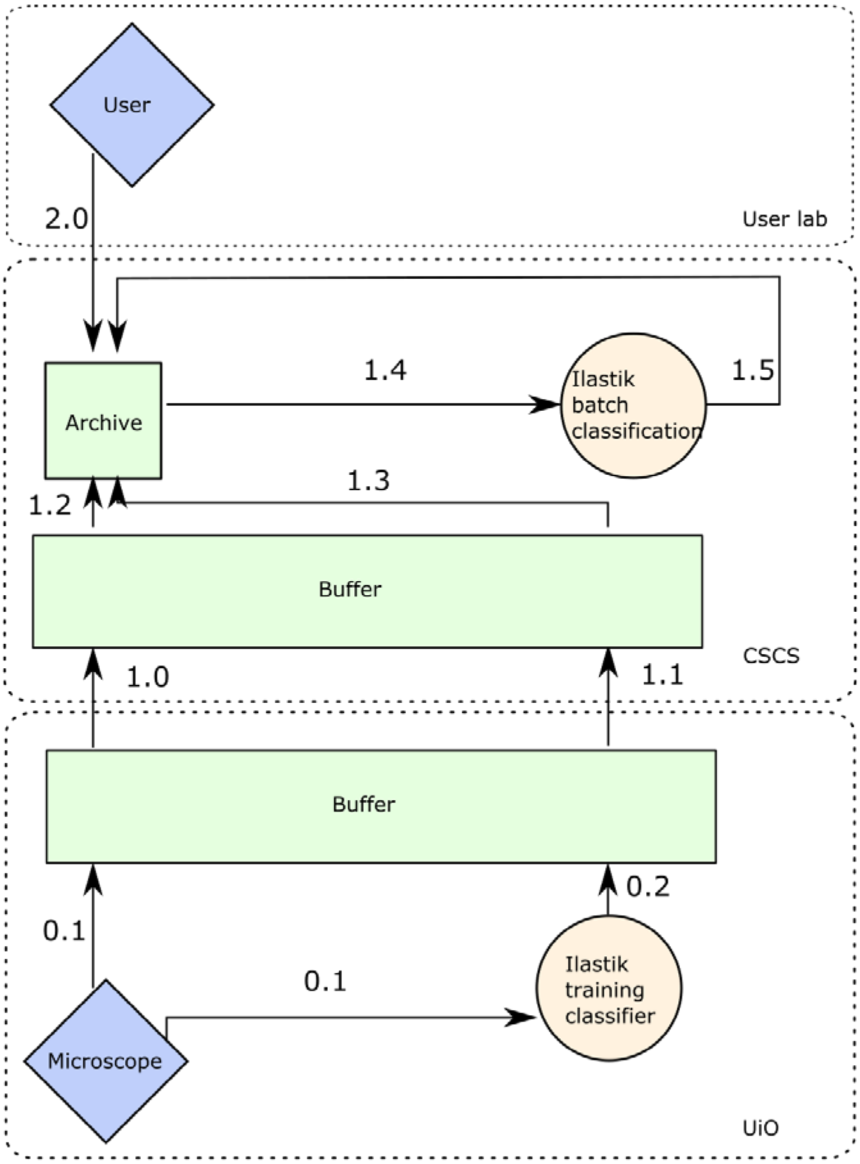

Figure 25. Schematic of the data flow for Ilastik Use Case 1

<u>Nodes @ UiO:</u>

- **Microscope** (data ingest):
  - o Data source: microscope glass
  - o Data characteristics:
    - tiled tiff uint8 between 300 Mbytes and 2 Gbytes.
    - Tiff could be LZW or JPG compressed
    - metadata embedded in the tiff file
- **Ilastik training classifier** (processing station)

- Data input: tiled tiff, uint8
- Description of steps
  - User load several images
  - User create *N* classes he wants to segment
  - User select in the image some pixels that are members of each of the *N* class
  - User train the classifier on these selected pixels and the classifier predicts the classes of the other pixels
  - The user interactively changes the class of some misclassified pixels and re-trains the classifier. This step is done enough times until the classifier is trained
  - The user saves the classifier.

- **Buffer** (Data repository) **- Transient**
  - Image data:
    - Format:tiled tiff , uint8
    - Tiff could be LZW or JPG compressed
    - Size: 60 Gbytes
    - Metadata: embedded in the tiff file (EXIF format)
  - Classifier data:
    - Format: HDF5
    - Size: Mbytes
    - Metadata: HDF5

Arrows @ UiO

- 0.1 :
  - General description of what data is transported : tiff files
- 0.2 :
  - General description of what data is transported : HDF5 files
- 1.1 :
  - General description of what data is transported : tiff files
  - Maximum required bandwidth
  - Average required bandwidth : transfer of full dataset should not take more than 30 minutes to CSCS
  - Data access patterns (request rate, transfer sizes) : 60 Gbytes
  - Interface requirements for attached entities : FTP or alternatives
- 1.2 :
  - General description of what data is transported : HDF5 files
  - Data access patterns (request rate, transfer sizes) : <Gbytes
  - Interface requirements for attached entities

Nodes @ CSCS:

- **Buffer** (Data repository) **- Transient**
  - Image data :
    - Format: tiled tiff, uint8
    - Size : 60Gbytes
    - Metadata: embedded in the tiff file (EXIF format)
  - Classifier data :
    - Format: Ilastik project (.ilp) (Python format)

        - Size: Mbytes

- **Archive** (Data repository) **- Permanent**
    - Image raw data
        - Format: tiled tiff
        - Size : 60Gbytes
        - Metadata: embedded in the tiff file (EXIF format)
        - Access: parallel access from all nodes (different nodes read different blocks of data), occurring at regular intervals, not necessarily continuous. Bandwidth: not necessarily real time access, but processing time. Interface: image service
    - Image processed data
        - Format: tiled tiff
        - Size : 60Gbytes
        - Metadata: embedded in the tiff file (EXIF format)
        - Interface: image service
    - Classifier data :
        - Format: HDF5
        - Size: up to a few Gigabytes, usually less than a Gigabyte
        - Access: load once at the start of the job
    - User batch processing definition data :
        - Format: JSON file
        - Size: Kbytes
        - Access: load once at the start of the job.
    -
    - Database required :
        - HBP KG Metadata database
        - Format: JSON-LD
        - Size: O(10Kbytes) per dataset.
        - Access: REST API
        - Other requirements/capabilities:
            - ACL support
            - Provenance support
            - Flexible indexing for user search.
            - Service must be deployed for access outside the CSCS network.
            - Sufficiently scalable to support high volume automated metadata registration from automated workflows.

- **Ilastik batch classification**
    - General description of data processing
        - Program loads classifier and image
        - Program applies classifier to image
        - Program saves the classified image
    - Data processing hardware architecture requirements
        - Intel CPUs, at least 8GB RAM per core
    - Required software stacks
        - Linux, i.e. CentOS6 or higher. Docker
    - Ratio of data processing rate versus data consumption and production rate

Arrows @ CSCS

- 1.4:
    - General description of what data is transported: tiff files
    - Interface requirements for attached entities: image service

- Reading
- 1.5:
    - General description of what data is transported: tiff files
    - Interface requirements for attached entities: image service
    - Writing

Nodes @ User lab:

- **User** (Data ingest)
    - Batch query data:
        - Format: JSON
        - Size: Kbytes

Arrows @ Userlab

- 2.0:
    - General description of what data is transported: JSON files

Workflow Description

The user is going to upload high resolution microscopic imaging data and has the objective to classify/segment objects in the microscopic image.

In this Use Case, the user will train the classifier locally on its machine. The optimised classifier will be transferred where the images are located. The images of the dataset will be processed in a batch manner on the server. The output of the batch processing (segmented images) will be stored together with the high resolution microscopic images and will be used in another Use Case.

Current solution(s)

The current solution is to do all this analysis process locally on a standard Linux/Windows machine.

Goal(s)

The goal of this Use Case is two-fold:

- have a mechanism to upload significant amount of data to an hbp server;
- run a program on the HBP server where the data is located to obtain processed data.

1 Johnson, G.A., Leergaard, T.B., Bjaalie, J.G. Unpublished data.

2 Johnson, G.A., Badea, A., Brandenburg, J., Cofer G., Fubara, B., Liu, S., Nissanov, J. *Waxholm Space: An image-based reference for coordinating mouse brain research*. Neuroimage, 2010;**53**:365-372.

3 Paxinos. G., Franklin, K. *Paxinos and Franklin's the mouse brain in stereotaxic coordinates*, 4th Edition. Amsterdam: Academic Press, 2013. 360 pp.

4 Cazemier, J.L., Clascá, F., Tiesinga, P.H.E. Connectomic analysis of brain networks: Novel techniques and future directions. *Front. Neuroanat*. 9 November 2016. DOI: 10.3389/fnana.2016.00110

5 Ji, S., Fakhry, A., Deng, H. Integrative analysis of the connectivity and gene expression atlases in the mouse brain. *Neuroimage.* 2013;**84**:245-253. DOI: 10.1016/j.neuroimage.2013.08.049

6 Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., *et al*. A mesoscale connectome of the mouse brain*. Nature.* 2014;**508**:207–214. doi:10.1038/nature13186.

7 Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature.* 2007;445:168-176. doi:10.1038/nature05453

8 Hinne, M., Meijers, A., Bakker, R., Tiesinga, P.H.E., Mørup, M., van Gerven, M.A.J. The missing link: Predicting connectomes from noisy and partially observed tract tracing data. *PLoS Computational Biology*. 2017 January 31 (corrected April 11). https://doi.org/10.1371/journal.pcbi.1005374